

Markov 过程导论

An Introduction to Markov Processes

Daniel W. Stroock 著

林正炎 张立新 苏中根 译

高等教育出版社
Higher Education Press

本书是一本很好的随机过程的入门书籍，主要介绍可数状态空间上的 Markov 过程的基本理论，内容包括 Doeblin 理论、一般的遍历理论、连续时间 Markov 过程和可逆 Markov 过程。书中每一章后面都提供了不同难易程度的练习题，并对一些富有挑战性的问题给出了有益的提示。为了便于阅读，最后一章概要地介绍了有关测度论的基础知识。

本书适用于具有较强数学背景的本科生，包括工程、经济、物理、生物等专业的学生，也可作为概率统计、数学等相关专业研究生的教材或参考书。


ISBN 978-7-04-022936-3

9 787040 229363 >

定价 32.00 元

学科类别：数学 / 概率

网址：academic.hep.com.cn



Markov过程导论

An Introduction to Markov Processes

■ Daniel W. Stroock 著

■ 林正炎 张立新 苏中根 译

高等教育出版社
Higher Education Press

International Press

图字: 01-2007-4659

Many thanks to Daniel W. Stroock for allowing Higher Education Press and International Press to publish and distribute the Chinese translation version of this book worldwide without a copyright charge.

感谢本书作者 Daniel W. Stroock 免费授予高等教育出版社、国际出版社本书的中文版出版权并在世界范围销售.

图书在版编目(CIP)数据

Markov 过程导论/(美)斯注克(Stroock, D.W.)著；林正炎，张立新，苏中根译. 北京：高等教育出版社，2007.12
(数学翻译丛书/丘成桐主编)
书名原文：An Introduction to Markov Processes
ISBN 978-7-04-022936-3

Ⅰ.M… Ⅱ.①斯…②林…③张…④苏… Ⅲ.马尔可夫过程-高等学校-教材 Ⅳ.O211.62

中国版本图书馆 CIP 数据核字(2007)第 163554 号

策划编辑	王丽萍	责任编辑	李华英	封面设计	王凌波
版式设计	马敬茹	责任校对	王　雨	责任印制	毛斯璐

出版发行	高等教育出版社	购书热线	010-58581118
社　　址	北京市西城区德外大街 4 号	免费咨询	800-810-0598
邮政编码	100011	网　　址	http://www.hep.edu.cn
总　　机	010-58581000		http://www.hep.com.cn
		网上订购	http://www.landraco.com
经　　销	蓝色畅想图书发行有限公司		http://www.landraco.com.cn
印　　刷	北京北苑印刷有限责任公司	畅想教育	http://www.widedu.com

开　　本	787×960　1/16	版　　次	2007 年 12 月第 1 版
印　　张	13	印　　次	2007 年 12 月第 1 次印刷
字　　数	200 000	定　　价	32.00 元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换。

物料号　22936-00

谨以此书献给长期一起工作的同事

Richard A. Holley

《数学翻译丛书》序

改革开放以后，国内大学逐渐与国外的大学增加交流. 无论到国外留学的学者及到中国访问的学者数量每年都有增长，对中国的科学现代化大有帮助. 但是在翻译外国文献方面的工作尚不多. 基本上所有中国的教科书都还是由本国教授撰写，有些已经比较陈旧，跟不上时代的发展了. 很多国家，例如俄罗斯、日本等，都大量翻译外文书来增加本国国民的阅读内容，对数学的研究大有裨益. 高等教育出版社和海外的国际出版社有鉴于此，开始计划做有系统的翻译，由王元院士领导，北京的晨兴数学中心和杭州的浙江大学数学科学研究中心共同组织这项工作. 参与的教授很多，有杨乐院士、刘克峰教授等. 我们希望这套翻译书能够使大学生有更多的角度来看数学，丰富他们的知识. 对这套丛书的出版，海外的出版公司如美国数学学会等多有帮助，我们谨此鸣谢.

丘成桐 (Shing-Tung Yau)

2005 年 1 月

序　　言

从某种意义上讲, 把本书看做介绍计算转移概率矩阵 $\boldsymbol{P}$ (即 $\boldsymbol{P}$ 中所有元素 $(\boldsymbol{P})_{ij}$ 非负并且 $\boldsymbol{P}$ 每一行元素和为 1) 的高阶幂或者 $\boldsymbol{R}(\boldsymbol{P}-\boldsymbol{I})$ 的指数幂 (其中 $\boldsymbol{R}$ 是一个具有非负元素的对角矩阵) 的内容的极大拓展是恰当的. 确实, 这是本书所做的事情. 然而, 对于有人如此轻视地对待 Markov 链和具有可数状态空间的 Markov 过程理论, 我和我的同仁们都深感愤怒; 我写这本书的主要目的正是为了向读者证明我们的不满是合理的.

我和持此信念的人之所以不同意把该理论只看做是矩阵理论的一部分, 理由是这样做忽视了概率贯穿始终所起的重要作用. 也就是说, 概率论提供一个模型, 启发并为我们处理矩阵所做的努力提供背景. 即 "转移概率矩阵" 这一术语使得构造一个矩阵的一系列相当特殊的假设变得有意义. 换句话说, 它意味着我们把矩阵元素 $(\boldsymbol{P})_{ij}$ 看做处于状态 i 的系统一步转移到状态 j 的概率. 进而, 按照这样的定义, 我们必须把 $\boldsymbol{P}^n$ 中的元素 $(\boldsymbol{P}^n)_{ij}$ 看成是 n 步的转移概率. 这样, 当 $n\to\infty$ 时, $\boldsymbol{P}^n$ 反映的是一个随机发展系统长时间的行为, 其中 $\boldsymbol{P}$ 描述一步转移. 正如我们将看到的那样, 这种解释将有助于我们对 $\lim\limits_{n\to\infty}(\boldsymbol{P}^n)_{ij}$ 的理解. 另外, 或许更重要的是, 概率扮演着连接数学和世界的桥梁角色. 事实上, 正是概率思想帮助人们为自然和社会科学中观察到的各种现象建立数学模型. 没有概率语言, 很难想象如何将这些现象和 $\boldsymbol{P}^n$ 联系在一起.

不管前一段结尾的观点如何, 本书是从一个数学家的视角来写的.

因而, 大多数情况下, 概率思想被用来诠释数学概念, 而不是为那些非数学现象提供数学解释. 这样做有两个理由. 第一个也是最重要的一个, 是我自己的背景. 尽管我偶尔也会尝试着去帮助那些从事各种应用的人们, 但我并没有因此而积累下很多实例可以在编写这样一本书的时候作为其中的内容. 事实上, 我的经验告诉我, 从事应用研究的那些人足以处理所遇到的常规问题, 他们求助像我这样的人, 实在是迫不得已. 因此, 他们的问题往往很难, 并且我能解决的那一小部分通常涉及超过本书范围的内容. 以目前这种方式写书的第二个理由是, 我认为材料本身就有足够的兴趣. 不管资助机构如何劝说, 数学或拟数学是一项有意义的科学事业, 并且我认为随机过程的全新引入应有一席之地, 它将毫无愧色地使数学达到至高境地.

在讲授由 MIT 数学系开设的随机过程导论几个学期之后, 我逐渐形成了这种观点. 听课学生一直是本科生和研究生的有趣混合, 其中不到一半的学生主修数学. 不过, 尽管缺少现代随机过程课程所必需的正规的数学训练, 至少当这门课目前在数学系为他们自己的研究生开设时, 大部分坚持听完这门课程的学生都有相当好的天赋和对数学的欣赏. 结果, 我发现, 没有现成的教材供选择. 一方面, 最容易的选择是 Karlin 的经典教材 "A First Course in Stochastic Processes", Karlin 编写的原著或者 Karlin 和 Taylor 的修订版 [4]. 他们的教材恰如其分地介绍了随机过程, 特别是可数状态空间上的 Markov 过程, 它那些简洁的证明, 如果不总是那么容易消化的话, 可以由大量的例子和练习来补充完整. 另一方面, 起初, 我担心采用 Karlin 和 Taylor 的教材会犯一个类似于采用 Feller 的著作作为本科生概率论入门的错误, 并且这种担心在我最初两次教学中出现过. 然而, 在使用并发现 Karlin 经典教材的两个改写版本以后, 我开始作出重大决定, 使用 Karlin 和 Taylor 的教材. 结果正如我所预料的那样: 我对这本教材的热情远远超过了学生.

为了使 Karlin 和 Taylor 的教材更适合于学生, 我开始补充一些注记, 试着重写其中的证明, 希望学生更容易懂些. 我的努力得到了回报, 学生对此反应积极. 事实上, 随着我的注记越来越多并开始淡化了原书的重要性, 我决定把它们变成现在这本书的样子, 尽管我意识到决定这样做可能有点愚蠢. 市场上有关这些内容的教材已经接近饱和. 而且, 其中一些书很受欢迎, 尽管据我了解, 它们受到欢迎并不总是和它们所包含的数学内容质量有关. 作了这样贬抑的评价, 我不便公开那

些书. 反之, 我想提及一下我非常喜欢的, 除了 Karlin 和 Taylor 的教材之外的书. Norris [5] 在介绍 Markov 过程方面是一本优秀教材, 同时让读者有很好的机会练习测度理论技巧. 当然, Norris 的教材仅仅适合那些具有测度理论技巧的学生. 另一方面, 对于掌握了那些技巧的学生而言, Norris 的教材让他们看到测度理论揉合在一起的魅力. 还有, Norris 给出了很多有趣的例子和练习, 用于说明如何应用. 本书包含了 [5] 中的大多数数学内容, 但证明却没有要求那么多测度论知识. 事实上, 尽管我系统地使用了第六章中解释的测度论术语 (Lebesgue 控制收敛定理、单调收敛定理等), 但这样做仅仅是为了让读者熟悉那些如果更深入地学习会遇到的专业术语. 说实话, 由于本书中的状态空间是可数的, 我所给出的 Lebesgue 理论的应用, 除一个外, 其余的都极为简单. 在 §6.2 中给出的这个例外包含了存在可列无穷多个相互独立随机变量的证明. 正如可能会出现的那样, 承认这样一族随机变量确实存在的读者, 除了术语和若干有关级数的明显结果外, 不需要参考第六章. 对程度更高一点的学生而言, 有关一般状态空间上 Markov 链的精彩内容可参考 Revuz [6].

本书的编排从目录上看一目了然. 第一章给出了一些基本的事实, 特别强调平面格点以及 d 维格点 $\mathbb{Z}^d$ 上最近邻随机游动的常返性和瞬时性. 第二章引入遍历性的研究, 这是联系第二章至第五章的主题. 第二章中所考虑的是 Markov 链 (即时间参数是离散的), 内容主要围绕 Doeblin 思想展开. 虽然 Doeblin 思想的适应性有局限性, 但与第三章和第四章的材料相比它有很大的优势, 它给出了 Markov 链收敛到其均衡分布的速度. 对 Doeblin 理论作了彻底的说明之后, 第三章研究当 Markov 链不满足 Doeblin 条件时的遍历性. 其主要结果由 (3.2.15) 式给出. 即使 (3.2.15) 的推导是完全初等的, 却毫无疑问是整本书最苛求的地方. 据我所知, (3.2.15) 的每一种证明都需要一些工作. 在想象的更简化的证明中, 这种工作隐藏在某些地方 (在测度论里, 如 [5] 和 [6]; 或者在算子论里, 如 [2]). 本书所给出的推导, 重新整理了 [4] 中基于 Feller 更新定理的证明, 仅要求读者透彻地理解极限的性质, 包括上极限、下极限以及极限的存在性. 在第四章, Markov 链由连续时间的 Markov 过程 (仍然具有可数状态空间) 所取代. 首先考虑速率有界的情形, 这样可能爆炸的问题就不会出现了. 接着, 考虑无界速率情形, 发展一些除有界性外, 能保证非爆炸的判别法则. 其余部分将第三章中有关 Markov 链的结论推广到连续时间情形. 除了更像附录而不

像全书的一个完整部分的第六章外, 本书到第五章为止. 第五章旨在当 Doeblin 理论完全失效或者产生相当差的估计时获得一些定量的结论, 这些结论令人回想起第二章的那些结论, 虽然它们没有第二章的结论强. 其中新的内容是假设 Markov 链或 Markov 过程是可逆的 (即转移概率在其平稳分布的 L^2 空间里是自伴随的), 并且内在发展机制是相伴的 Dirichlet 型. 最后一节, Dirichlet 型方法的功效在 Metropolis(模拟退火) 算法的分析上得到检验. 最后, 正如前面所说的, 第六章是一个附录, 回顾了有关测度和积分的 Lebesgue 理论的一些思想和术语. 相当大的一部分是前面提到过的、§6.2.1 中的构造.

最后, 按照惯例, 应该对那些直接或者间接地为本书做出贡献的人表示感谢. 主要直接贡献者是学生, 他们试用着各种各样的临时修订的版本. 我特别感谢 Adela Popescu, 他仔细阅读了本书, 指出很多小错误和几个严重错误, 现都已被删除或改正了. 感谢甚至找到那些间接做出贡献的人是比较困难的. 事实上, 包括所有那些已故和健在的人们, 从他们那儿我受到教育. 我不打算列出哪怕一小部分人的名单, 告诉你他们是谁. 不过, 有一个人, 在十多年的时间里, 一直耐心地教我欣赏本书所写的内容. 他就是我要将这本书献给的那位, Richard A. Holley, 一位真正的概率学家. 对 Dick 来说, 关于概率现象, 直觉通常先于数学上严格的理解. 这样说不应该使人怀疑 Dick 作为一名严谨数学家的能力. 相反地, 他关于概率理论的直观理解不仅增强了他自己那令人敬畏的数学功力, 也把我和其他一些人从盲目地陷入不完善的推理中解救出来. 正如所有和他一起共事的人都知道的那样, 重新思量着你所说的, 甚至一些激烈的抨击, Dick 只是轻轻地说: "我倒不这样认为."

除了他数学上的非凡能力, Dick 的众多学生中人人都会证明他的慷慨大度. 我不是他的学生, 但我是他的同事, 可以向你保证, 他的慷慨大度绝不仅仅限于他的学生.

Daniel W. Stroock, 2004 年 8 月

目　录

序言 · i

第一章　随机游动 —— 一个好的切入点 · · · · · · · · · · 1

1.1　$\mathbb{Z}$ 上最近邻随机游动 · · · · · · · · · · · · · · 1

1.1.1.　n 时刻的分布 · 2

1.1.2.　利用反射原理研究通过次数 · · · · · · · · · · · · 3

1.1.3.　若干相关的计算 · · · · · · · · · · · · · · · · · · 5

1.1.4.　首次返回的时刻 · · · · · · · · · · · · · · · · · · 7

1.1.5.　利用泛函方程研究通过次数 · · · · · · · · · · · · 8

1.2　随机游动的常返性 · · · · · · · · · · · · · · · · · 9

1.2.1.　$\mathbb{Z}^d$上的随机游动 · · · · · · · · · · · · · · · · · 9

1.2.2.　一个初等的常返性判别法则 · · · · · · · · · · · 10

1.2.3.　$\mathbb{Z}^2$上对称随机游动的常返性 · · · · · · · · · · 12

1.2.4.　$\mathbb{Z}^3$上的瞬时性 · · · · · · · · · · · · · · · · · · 14

1.3　习题 · 17

第二章　Markov 链的 Doeblin 理论 · · · · · · · · · · · 25

2.1　概论 · 25

2.1.1.　Markov 链的存在性 · · · · · · · · · · · · · · · · · 26

2.1.2.　转移概率和概率向量 · · · · · · · · · · · · · · · · 27

2.1.3. 转移概率和转移函数 …… 28
2.1.4. Markov 性 …… 30
2.2 Doeblin 理论 …… 30
2.2.1. Doeblin 基本定理 …… 30
2.2.2. 两个推广 …… 33
2.3 遍历理论要素 …… 35
2.3.1. 平均遍历定理 …… 35
2.3.2. 返回次数 …… 37
2.3.3. π 的确定 …… 41
2.4 习题 …… 43

第三章 Markov 链的遍历理论 (续) …… **48**
3.1 状态的分类 …… 49
3.1.1. 分类、常返性和瞬时性 …… 49
3.1.2. 常返性和瞬时性的判别法则 …… 52
3.1.3. 周期性 …… 55
3.2 没有 Doeblin 条件的遍历理论 …… 57
3.2.1. 矩阵的收敛性 …… 57
3.2.2. Abel 收敛性 …… 59
3.2.3. 平稳分布的结构 …… 62
3.2.4. 一个小的改进 …… 64
3.2.5. 平均遍历定理 (续) …… 66
3.2.6. 非周期情形的一个改进 …… 68
3.2.7. 周期性结构 …… 71
3.3 习题 …… 73

第四章 连续时间 Markov 过程 …… **82**
4.1 Poisson 过程 …… 82
4.1.1. 简单 Poisson 过程 …… 82
4.1.2. $\mathbb{Z}^d$上 的复合 Poisson 过程 …… 85
4.2 带有界速率的 Markov 过程 …… 88
4.2.1. 基本结构 …… 88
4.2.2. Markov 性 …… 91

4.2.3. Q-矩阵和 Kolmogorov 向后方程 ······ 93
4.2.4. Kolmogorov 向前方程 ······ 94
4.2.5. 解 Kolmogorov 方程 ······ 94
4.2.6. 具有无穷小特征的 Markov 过程 ······ 96
4.3 无界速率 ······ 97
4.3.1. 爆炸 ······ 97
4.3.2. 非爆炸或爆炸的准则 ······ 100
4.3.3. 当爆炸发生时做什么 ······ 103
4.4 遍历性质 ······ 104
4.4.1. 状态的分类 ······ 104
4.4.2. 平稳测度与极限定理 ······ 107
4.4.3. 解释 $\hat{\pi}_{ii}$ ······ 110
4.5 习题 ······ 111

第五章 可逆 Markov 过程 ······ 117
5.1 可逆 Markov 链 ······ 118
5.1.1. 从不变性到可逆性 ······ 118
5.1.2. 二次平均度量 ······ 118
5.1.3. 谱隙 ······ 121
5.1.4. 可逆性和周期性 ······ 123
5.1.5. 与变差收敛的关系 ······ 124
5.2 Dirichlet 型和 β 的估计 ······ 126
5.2.1. Dirichlet 型和 Poincaré 不等式 ······ 126
5.2.2. β_+ 的估计 ······ 129
5.2.3. β_- 的估计 ······ 130
5.3 连续时间可逆 Markov 过程 ······ 132
5.3.1. 可逆性准则 ······ 132
5.3.2. 有界速率时 $L^2(\hat{\pi})$ 中的收敛性 ······ 133
5.3.3. 一般情形下 $L^2(\hat{\pi})$-收敛速度 ······ 134
5.3.4. 估计 λ ······ 137
5.4 Gibbs 态和 Glauber 动力系统 ······ 138
5.4.1. 框架 ······ 138
5.4.2. Dirichlet 型 ······ 140

5.5 模拟退火 · · · · · · 143
5.5.1. 算法 · · · · · · 143
5.5.2. 转移概率的构造 · · · · · · 144
5.5.3. Markov 过程的描述 · · · · · · 146
5.5.4. 冷却方案的选取 · · · · · · 147
5.5.5. 小的改进 · · · · · · 149
5.6 习题 · · · · · · 151

第六章 测度理论简介 · · · · · · **159**
6.1 Lebesgue 测度理论 · · · · · · 159
6.1.1. 测度空间 · · · · · · 159
6.1.2. 关于可数可加性的一些结论 · · · · · · 161
6.1.3. 生成 σ-代数 · · · · · · 162
6.1.4. 可测函数 · · · · · · 163
6.1.5. Lebesgue 积分 · · · · · · 164
6.1.6. Lebesgue 积分的稳定性 · · · · · · 166
6.1.7. 可数空间上的 Lebesgue 积分 · · · · · · 168
6.1.8. Fubini 定理 · · · · · · 170
6.2 概率建模 · · · · · · 172
6.2.1. 无穷多次投掷均匀硬币的模型 · · · · · · 173
6.3 独立随机变量 · · · · · · 178
6.3.1. 独立随机变量族的存在性 · · · · · · 179
6.4 条件概率和条件期望 · · · · · · 181
6.4.1. 关于随机变量的条件运算 · · · · · · 182

符号 · · · · · · **184**

参考文献 · · · · · · **186**

索引 · · · · · · **187**

第一章　随机游动
—— 一个好的切入点

本章旨在讨论 Markov 过程的一些例子，它们甚至可以先于 "Markov 过程" 被人们所理解. 事实上, 任何学过概率论的人都会发现, 这些过程都源于基本的 "投掷硬币" 的研究.

1.1　$\mathbb{Z}$ 上最近邻随机游动

令 p 是开区间 $(0,1)$ 上的一个数, 假设[1] $\{B_n : n \in \mathbb{Z}^+\}$ 是一列取 $\{-1,1\}$ 值、同分布的 *Bernoulli* 随机变量 [2], 其中取 1 的概率是 p. 即对任何 $n \in \mathbb{Z}^+$ 和任意 $E \equiv (\epsilon_1, \cdots, \epsilon_n) \in \{-1,1\}^n$,

$$\begin{aligned} &\mathrm{P}(B_1 = \epsilon_1, \cdots, B_n = \epsilon_n) = p^{N(E)} q^{n-N(E)}, \quad \text{其中 } q \equiv 1-p, \\ &N(E) = \#\{m : \epsilon_m = 1\} = \frac{n + S_n(E)}{2}, \quad \text{其中 } S_n(E) \equiv \sum_1^n \epsilon_m. \end{aligned} \tag{1.1.1}$$

下面, 令

$$X_0 = 0, \quad X_n = \sum_{m=1}^n B_m, \qquad n \in \mathbb{Z}^+. \tag{1.1.2}$$

$\{B_n : n \in \mathbb{Z}^+\}$ 的存在性可参看 §6.2.1.

[1] $\mathbb{Z}$ 代表整数集, $\mathbb{N}$ 和 $\mathbb{Z}^+$ 分别表示非负整数集和正整数集.

[2] 由于历史的原因, 相互独立、只取两个值的随机变量往往称为 Bernoulli 随机变量.

上述随机变量族 $\{X_n : n \in \mathbb{N}\}$ 通常被称为 $\mathbb{Z}$上的**最近邻随机游动**. 最近邻随机游动是 Markov 过程的一个例子, 但是我们刚刚所给出的描述是初等概率论中所有的; 概率论和本课程 —— 研究随机过程的课程 —— 不同. 也就是说, 在随机过程的研究中, 应该强调随机变量族的动态方面. 因此, 一个具有随机过程特色的描述可以用

$$\mathrm{P}(X_0 = 0) = 1,$$
$$\mathrm{P}(X_n - X_{n-1} = \epsilon | X_0, \cdots, X_{n-1}) = \begin{cases} p, & 若\ \epsilon = 1; \\ q, & 若\ \epsilon = -1 \end{cases} \tag{1.1.3}$$

取代 (1.1.2) 式, 其中 $\mathrm{P}(X_n - X_{n-1} = \epsilon | X_0, \cdots, X_{n-1})$ 表示在给定 $\sigma(\{X_0, \cdots, X_{n-1}\})$ 后, $X_n - X_{n-1} = \epsilon$ 的**条件概率** (参看 §6.4.1). 注意, (1.1.3) 式比 (1.1.2) 式更具动态性. 特别地, 它指出该过程在 $n = 0$ 时刻从 0 开始, 在每个时刻 $n \in \mathbb{Z}^+$, 以概率 p 前进一步或者以概率 q 后退一步, 并且和时刻 n 以前所处的位置无关.

1.1.1. n 时刻的分布: 在这一小节, 我们提出计算 $\mathrm{P}(X_n = m)$ 的两种方法. 第一种方法基于 (1.1.2) 式. 从 (1.1.2) 式容易得到 $\mathrm{P}(|X_n| \leqslant n) = 1$. 另外, 显然有

$$n\ 为奇数 \implies \mathrm{P}(X_n\ 为奇数) = 1 \quad 和$$
$$n\ 为偶数 \implies \mathrm{P}(X_n\ 为偶数) = 1.$$

最后, 给定和 n 具有相同奇偶性的 $m \in \{-n, \cdots, n\}$ 以及 $E = (\epsilon_1, \cdots, \epsilon_n) \in \{-1, 1\}^n$ (参看 (1.1.1)) 使得 $S_n(E) = m$, $N(E) = \dfrac{n+m}{2}$, 因此

$$\mathrm{P}(B_1 = \epsilon_1, \cdots, B_n = \epsilon_n) = p^{\frac{n+m}{2}} q^{\frac{n-m}{2}}.$$

这样, 记 $\binom{\ell}{k} \equiv \dfrac{\ell!}{k!(\ell-k)!}$ 表示**二项系数**"从 ℓ 中选择 k 个", 由于存在 $\begin{pmatrix} n \\ \frac{m+n}{2} \end{pmatrix}$ 个这样的 E, 可知如果 $m \in \mathbb{Z}$, $|m| \leqslant n$ 且 m 和 n 具有相同的奇偶性, 那么

$$\mathrm{P}(X_n = m) = \begin{pmatrix} n \\ \frac{m+n}{2} \end{pmatrix} p^{\frac{n+m}{2}} q^{\frac{n-m}{2}}, \tag{1.1.4}$$

否则为 0.

第二种计算这个概率的方法将基于 (1.1.3) 式的动态描述. 为此, 引进记号 $(P^n)_m \equiv \mathrm{P}(X_n = m)$. 显然, $(P^0)_m = \delta_{0,m}$, 其中 $\delta_{k,\ell}$ 是

Kronecker 符号, 当 $k=\ell$ 时取 1, 其他情形取 0. 进而, 由 (1.1.3) 式知, $\mathrm{P}(X_n=m)$ 等于

$$\begin{aligned}&\mathrm{P}(X_{n-1}=m-1,X_n=m)+\mathrm{P}(X_{n-1}=m+1,X_n=m)\\&=p\mathrm{P}(X_{n-1}=m-1)+q\mathrm{P}(X_{n-1}=m+1).\end{aligned}$$

即

$$(P^0)_m=\delta_{0,m},\quad (P^n)_m=p(P^{n-1})_{m-1}+q(P^{n-1})_{m+1}.\qquad(1.1.5)$$

显然, (1.1.5) 式给出了一个计算 $(P^n)_m$ 的完整的递推方法, 尽管这一方法是用隐式表达的, 并且很容易验证 (1.1.4) 式给出的数值满足这个递推式. 反过来, 可以用 (1.1.5) 式加上对 n 的递归得出: 除非对某个 $0\leqslant\ell\leqslant n$, $m=2\ell-n$, 否则有 $(P^0)_m=0$, 并且若记 $(C^n)_\ell\equiv p^{-\ell}q^{n-\ell}(P^n)_{2\ell-n}$, 则 $(C^n)_\ell=(C^{n-1})_{\ell-1}+(C^{n-1})_\ell$. 换句话说, 系数 $\{(C^n)_\ell:n\in\mathbb{N},0\leqslant\ell\leqslant n\}$ 由 Pascal 三角形给出, 因此正好是二项系数, 从而 (1.1.4) 成立.

1.1.2. 利用反射原理研究通过次数: 比 §1.1.1 中的计算更有挑战性的是确定某一点 $a\in\mathbb{Z}$ 的首达时的分布. 即, 给定 $a\in\mathbb{Z}\backslash\{0\}$, 令[3]

$$\zeta_a=\inf\{n\geqslant 1:X_n=a\}(\equiv\infty,\ \text{如果对任意的 } n\geqslant 1,\ X_n\neq a).\qquad(1.1.6)$$

那么 ζ_a 是 a 点的首达时, 这里我们的目标是找出它的分布. 等价地, 我们要求出 $\mathrm{P}(\zeta_a=n)$ 的表达式. 显然, 由 §1.1.1, 我们只需要考虑满足 $n\geqslant|a|$ 并且和 a 具有相同奇偶性的那些 n.

我们给出这个问题的两种解法. 这里基于 (1.1.2) 式; 在 §1.1.5 中将基于 (1.1.3) 式. 为了叙述基于 (1.1.2) 式的方法, 假设 $a\in\mathbb{Z}^+$, 同时假设 $n\in\mathbb{Z}^+$ 和 a 具有相同的奇偶性. 首先, 注意到

$$\mathrm{P}(\zeta_a=n)=\mathrm{P}(X_n=a,\zeta_a>n-1)=p\,\mathrm{P}(\zeta_a>n-1,X_{n-1}=a-1).$$

因此, 我们只要计算 $\mathrm{P}(\zeta_a>n-1,X_{n-1}=a-1)$ 即可. 为此, 注意到对任意满足 $S_{n-1}(E)=a-1$ 的 $E\in\{-1,1\}^{n-1}$, 事件 $\{(B_1,\cdots,B_{n-1})=E\}$ 具有概率 $p^{\frac{n+a}{2}-1}q^{\frac{n-a}{2}}$. 因此

$$\mathrm{P}(\zeta_a=n)=\mathcal{N}(n,a)p^{\frac{n+a}{2}}q^{\frac{n-a}{2}},\qquad(*)$$

[3] 正如下面所示的那样, 我们取空集的下确界为 $+\infty$.

其中 $\mathcal{N}(n,a)$ 是 $\{-1,1\}^{n-1}$ 中满足下列条件的 E 的个数: 对 $0 \leqslant \ell \leqslant n-1$, $S_\ell(E) \leqslant a-1$ 且 $S_{n-1}(E)=a-1$. 这就是说, 最终归结为 $\mathcal{N}(n,a)$ 的计算. 另一方面 $\mathcal{N}(n,a)=\begin{pmatrix} n-1 \\ \frac{n+a}{2}-1 \end{pmatrix}-\mathcal{N}'(n,a)$, 其中 $\mathcal{N}'(n,a)$ 是 $\{-1,1\}^{n-1}$ 中使得 $S_{n-1}(E)=a-1$ 和 $S_\ell(E) \geqslant a$ (对某个 $\ell \leqslant n-1$) 的 E 的个数, 我们只需要计算 $\mathcal{N}'(n,a)$. 为此, 我们使用被称为**反射原理**的漂亮论证方法. 考虑所有具有性质 $S_0=0$, $S_m-S_{m-1} \in \{-1,1\}$ (对 $1 \leqslant m \leqslant n-1$), 且 $S_m \geqslant a$ (对某个 $1 \leqslant m \leqslant n-1$) 的路径 $(S_0,\cdots,S_{n-1}) \in \mathbb{Z}^n$ 所组成的集合 $P(n,a)$. 显然, $\mathcal{N}'(n,a)$ 是 $P(n,a)$ 中所有满足 $S_{n-1}=a-1$ 的 $(S_0,\cdots,S_{n-1})$ 组成的集合 $L(n,a)$ 中路径的数目. 应用反射原理, 我们将证明集合 $L(n,a)$ 和 $U(n,a)$ 所含的元素个数相同, 其中 $U(n,a)$ 是 $P(n,a)$ 中所有满足 $S_{n-1}=a+1$ 的路径 $(S_0,\cdots,S_{n-1})$ 组成的集合. 因为 $(S_0,\cdots,S_{n-1}) \in U(n,a)$ 当且仅当 $S_0=0$, $S_m-S_{m-1} \in \{-1,1\}$ (对所有 $1 \leqslant m \leqslant n-1$), 以及 $S_{n-1}=a+1$, 所以我们已经知道如何去计算它们: 共有 $\begin{pmatrix} n-1 \\ \frac{n+a}{2} \end{pmatrix}$ 条路径. 因此, 余下的是应用反射原理. 为此, 对给定的 $\boldsymbol{S}=(S_0,\cdots,S_{n-1}) \in P(n,a)$, 令 $\ell(\boldsymbol{S})$ 为使得 $S_k \geqslant a$ 的 $0 \leqslant k \leqslant n-1$ 中的最小的一个, 并定义 $\boldsymbol{S}$ 的**反射路径** $\mathcal{R}(\boldsymbol{S})=(\hat{S}_0,\cdots,\hat{S}_{n-1})$, 使得 $\hat{S}_m=S_m$ (若 $0 \leqslant m \leqslant \ell(\boldsymbol{S})$), $\hat{S}_k=2a-S_k$ (若 $\ell(\boldsymbol{S})<m \leqslant n-1$). 显然 $\mathcal{R}$ 将 $L(n,a)$ 映射到 $U(n,a)$, 将 $U(n,a)$ 映射到 $L(n,a)$. 另外, $\mathcal{R}$ 是幂等的: 与其自身复合是恒等映射. 所以, 作为一个从 $L(n,a)$ 到 $U(n,a)$ 的映射, $\mathcal{R}$ 一定是一一对应, 而且是满射. 因此, $L(n,a)$ 和 $U(n,a)$ 所包含的路径个数相等.

我们已经证明了 $\mathcal{N}'(n,a)=\begin{pmatrix} n-1 \\ \frac{n+a}{2} \end{pmatrix}$, 因此有

$$\mathcal{N}(n,a)=\begin{pmatrix} n-1 \\ \frac{n+a}{2}-1 \end{pmatrix}-\begin{pmatrix} n-1 \\ \frac{n+a}{2} \end{pmatrix}.$$

最后, 将它代入 $(*)$ 式可得

$$\mathrm{P}(\zeta_a=n)=\left[\begin{pmatrix} n-1 \\ \frac{n+a}{2}-1 \end{pmatrix}-\begin{pmatrix} n-1 \\ \frac{n+a}{2} \end{pmatrix}\right] p^{\frac{n+a}{2}} q^{\frac{n-a}{2}},$$

化简上述公式可以写成

$$\mathrm{P}(\zeta_a=n)=\frac{a}{n}\begin{pmatrix} n \\ \frac{n+a}{2} \end{pmatrix} p^{\frac{n+a}{2}} q^{\frac{n-a}{2}}=\frac{a}{n}\mathrm{P}(X_n=a).$$

$a<0$ 情形的计算可以重复刚刚的讨论给出, 或者将 p 和 q 互换, 再将前面的结果用到 $-a$ 上. 不管哪种方法, 一般的表达式为当 $n \geqslant |a|$ 且和 a 具有相同的奇偶性时, 成立

$$a \neq 0 \implies \mathrm{P}(\zeta_a = n) = \frac{|a|}{n}\binom{n}{\frac{n+a}{2}} p^{\frac{n+a}{2}} q^{\frac{n-a}{2}} = \frac{|a|}{n}\mathrm{P}(X_n = a), \quad (1.1.7)$$

其他情况下为 0.

1.1.3. 若干相关的计算: 尽管公式 (1.1.7) 很漂亮, 但不是十分清晰. 特别地, 完全不清楚如何用这个公式去判断 $\mathrm{P}(\zeta_a < \infty) = 1$ 是否成立. 为了计算它, 给定 $a > 0$, 并记 $\zeta_a = f_a(B_1, \cdots, B_n, \cdots)$, 其中 f_a 是一个函数, 将 $\{-1,1\}^{\mathbb{Z}^+}$ 映射到 $\mathbb{Z}^+ \bigcup \{\infty\}$, 使得对每个 $n \in \mathbb{N}$, 有

$$f_a(\epsilon_1, \cdots, \epsilon_n, \cdots) > n \iff \sum_{\ell=1}^{m} \epsilon_\ell < a, \quad 1 \leqslant m \leqslant n.$$

既然事件 $\{\zeta_a = m\}$ 仅仅依赖于 $(B_1, \cdots, B_m)$ 且

$$\zeta_a = m \implies \zeta_{a+1} = m + \zeta_1 \circ \Sigma^m, \quad (1.1.8)$$

其中 $\zeta_1 \circ \Sigma^m \equiv f_1(B_{m+1}, \cdots, B_{m+n}, \cdots)$, 那么 $\{\zeta_a = m, \zeta_{a+1} < \infty\} = \{\zeta_a = m\} \bigcap \{\zeta_1 \circ \Sigma^m < \infty\}$, 并且 $\{\zeta_a = m\}$ 和 $\{\zeta_1 \circ \Sigma^m < \infty\}$ 是相互独立的. 特别地, 因为 $(B_{m+1}, \cdots, B_{m+n}, \cdots)$ 和 $(B_1, \cdots, B_n, \cdots)$ 具有相同的分布且 $\zeta_1 \circ \Sigma^m$ 和 ζ_1 也具有相同的分布, 上述结论导出

$$\begin{aligned}
\mathrm{P}(\zeta_{a+1} < \infty) &= \sum_{m=1}^{\infty} \mathrm{P}(\zeta_a = m, \zeta_{a+1} < \infty) \\
&= \sum_{m=1}^{\infty} \mathrm{P}(\zeta_a = m)\mathrm{P}(\zeta_1 \circ \Sigma^m < \infty) \\
&= \mathrm{P}(\zeta_1 < \infty) \sum_{m=1}^{\infty} \mathrm{P}(\zeta_a = m) \\
&= \mathrm{P}(\zeta_1 < \infty)\mathrm{P}(\zeta_a < \infty).
\end{aligned}$$

当 $a<0$ 时, 用 -1 代替 1 同理可得. 换句话说, 我们已经证明了

$$\mathrm{P}(\zeta_a < \infty) = \mathrm{P}(\zeta_{\mathrm{sgn}(a)} < \infty)^{|a|}, \quad a \in \mathbb{Z} \backslash \{0\}, \quad (1.1.9)$$

其中 $\mathrm{sgn}(a)$ 是 a 的符号, 根据 $a > 0$ 或 $a < 0$ 分别取 1 或 -1. 特别地, 这证明了对所有的 $a \in \mathbb{Z}^+$, $\mathrm{P}(\zeta_1 < \infty) = 1 \implies \mathrm{P}(\zeta_a < \infty) = 1$, $\mathrm{P}(\zeta_{-1} < \infty) = 1 \implies \mathrm{P}(\zeta_{-a} < \infty) = 1$.

综上所述, 我们只需要考虑 $\mathrm{P}(\zeta_1 < \infty)$. 而由单调收敛定理, 定理 6.1.9, 可得

$$\mathrm{P}(\zeta_1 < \infty) = \lim_{s \nearrow 1} \mathrm{E}[s^{\zeta_1}] = \lim_{s \nearrow 1} \sum_{n=1}^{\infty} s^{2n-1} \mathrm{P}(\zeta_1 = 2n-1).$$

取 $a = 1$ 并利用 (1.1.7) 式, 可知

$$\mathrm{P}(\zeta_1 = 2n-1) = \frac{1}{2n-1}\binom{2n-1}{n} p^n q^{n-1}.$$

注意到

$$\begin{aligned}\frac{1}{2n-1}\binom{2n-1}{n} &= \frac{(2(n-1))!}{n!(n-1)!} = \frac{2^{n-1}}{n!}\prod_{m=1}^{n-1}(2m-1)\\ &= \frac{4^{n-1}}{n!}\prod_{m=1}^{n-1}\left(m-\frac{1}{2}\right) = (-1)^{n-1}\frac{4^n}{2}\binom{\frac{1}{2}}{n},\end{aligned}$$

其中[4], 对任何 $\alpha \in \mathbb{R}$,

$$\binom{\alpha}{n} \equiv \begin{cases} 1, & \text{若 } n = 0; \\ \dfrac{1}{n!}\displaystyle\prod_{m=0}^{n-1}(\alpha - m), & \text{若 } n \in \mathbb{Z}^+ \end{cases}$$

是 $(1+x)^\alpha$ 在 $x = 0$ 点的 Taylor 展开式中 x^n 的广义二项系数. 因此,

$$\sum_{n=1}^{\infty} s^{2n-1}\mathrm{P}(\zeta_1 = 2n-1) = -\frac{1}{2qs}\sum_{n=1}^{\infty}\binom{\frac{1}{2}}{n}(-4pqs^2)^n = \frac{1-\sqrt{1-4pqs^2}}{2qs}.$$

并由此得

$$\mathrm{E}[s^{\zeta_1}] = \frac{1-\sqrt{1-4pqs^2}}{2qs}, \quad |s| < 1. \tag{1.1.10}$$

当然, 根据对称性, 交换 p 和 q 可得

$$\mathrm{E}[s^{\zeta_{-1}}] = \frac{1-\sqrt{1-4pqs^2}}{2ps}, \quad |s| < 1. \tag{1.1.11}$$

在 (1.1.10) 式中令 $s \nearrow 1$, 并注意到 $1 - 4pq = (p+q)^2 - 4pq = (p-q)^2$, 我们有[5]

$$\lim_{s \nearrow 1} \mathrm{E}[s^{\zeta_1}] = \frac{1-|p-q|}{2q} = \frac{p \wedge q}{q},$$

---

[4] 前面我们已经约定, 若 $\ell < k$, $\prod_{j=k}^{\ell} a_j = 1$.

[5] $a \wedge b$ 表示 $\min\{a, b\}, a, b \in \mathbb{R}$.

并由此得

$$P(\zeta_1 < \infty) = \begin{cases} 1, & \text{若 } p \geqslant q; \\ \dfrac{p}{q}, & \text{若 } p < q. \end{cases}$$

当然, $P(\zeta_{-1} < \infty)$ 由同样的公式给出, 只要将 p 和 q 进行交换即可. 这样,

$$P(\zeta_a < \infty) = \begin{cases} 1, & \text{若 } a \in \mathbb{Z}^+, p \geqslant q \text{ 或 } -a \in \mathbb{Z}^+, p \leqslant q; \\ \left(\dfrac{p}{q}\right)^a, & \text{若 } a \in \mathbb{Z}^+, p < q \text{ 或 } -a \in \mathbb{Z}^+, p > q. \end{cases} \tag{1.1.12}$$

1.1.4. 首次返回的时刻: 在克服了许多困难获得 (1.1.12) 式之后, 希望能从它推导出 $\mathbb{Z}$ 上最近邻随机游动*常返性* 的一个著名结论. 令

$$\rho_0 \equiv \inf\{n \geqslant 1 : X_n = 0\} (\equiv \infty, \text{ 如果对所有 } n \geqslant 1, X_n \neq 0)$$

是首次返回到 0 的时刻. 那么, 根据和 (1.1.9) 式完全一样的推理可得 $P(X_1 = 1, \rho_0 < \infty) = pP(\zeta_{-1} < \infty)$ 和 $P(X_1 = -1, \rho_0 < \infty) = qP(\zeta_1 < \infty)$, 因此利用 (1.1.12) 式得

$$P(\rho_0 < \infty) = 2(p \wedge q). \tag{1.1.13}$$

换句话说, *随机游动 $\{X_n : n \geqslant 0\}$ 以概率 1 回到 0 当且仅当它是对称的*, 即 $p = \dfrac{1}{2}$.

将上面推理稍微加强一点, 可得 $P(X_1 = 1, \rho_0 = 2n) = pP(\zeta_{-1} = 2n-1)$ 和 $P(X_1 = -1, \rho_0 = 2n) = qP(\zeta_1 = 2n-1)$, 因此利用 (1.1.10) 式和 (1.1.11) 式, 得

$$E[s^{\rho_0}] = 1 - \sqrt{1 - 4pqs^2}, \quad |s| < 1. \tag{1.1.14}$$

因此

$$E[\rho_0 s^{\rho_0}] = s\frac{d}{ds}E[s^{\rho_0}] = \frac{4pqs^2}{\sqrt{1-4pqs^2}}, \quad |s| < 1,$$

并且因为[6] 当 $s \nearrow 1$ 时, $E[\rho_0 s^{\rho_0}] \nearrow E[\rho_0, \rho_0 < \infty]$, 所以

$$E[\rho_0, \rho_0 < \infty] = \frac{4pq}{|p-q|},$$

[6] 当 X 是一个随机变量, A 是一个事件时, 我们常用 $E[X, A]$ 表示 $E[X1_A]$.

它和 (1.1.13) 式相结合给出[7]

$$\mathrm{E}[\rho_0|\rho_0<\infty]=\frac{2p\vee q}{|p-q|}=1+\frac{1}{|p-q|}. \qquad (1.1.15)$$

上面所得出的结论给出了 $\mathbb{Z}$ 上最近邻随机游动性质的一个重要刻画. 首先, 当随机游动对称时, 它以概率 1 回到 0, 但所需要的平均时间是无穷大. 其次, 当随机游动不对称时, 它将以正概率不返回. 另一方面, 在不对称情况下, 样本轨迹具有很有趣的性质. 也就是说, (1.1.13) 式和 (1.1.15) 式表明它们要么从不返回, 要么迅速地返回.

1.1.5. 利用泛函方程研究通过次数: 我们以较少计算量再次推导 (1.1.10) 式, 并以此结束最近邻随机游动通过次数的讨论. 为此, 对 $a\in\mathbb{Z}\setminus\{0\}$ 和 $s\in(-1,1)$, 令 $u_a(s)=\mathrm{E}[s^{\zeta_a}]$. 给定 $a\in\mathbb{Z}^+$, 使用 §1.1.3 的思想, 特别是 (1.1.8) 式, 得到

$$\begin{aligned}u_{a+1}(s)&=\sum_{m=1}^{\infty}s^m\mathrm{E}[s^{\zeta_1\circ\Sigma^m},\zeta_a=m]=\sum_{m=1}^{\infty}s^m\mathrm{P}(\zeta_a=m)\mathrm{E}[s^{\zeta_1\circ\Sigma^m}]\\&=\sum_{m=1}^{\infty}s^m\mathrm{P}(\zeta_a=m)u_1(s)=u_a(s)u_1(s).\end{aligned}$$

类似地, 如果 $-a\in\mathbb{Z}^+$, 那么 $u_{a-1}(s)=u_a(s)u_{-1}(s)$. 因此

$$u_a(s)=u_{\mathrm{sgn}(a)}(s)^{|a|},\quad a\in\mathbb{Z}\setminus\{0\}\text{ 且 }|s|<1. \qquad (1.1.16)$$

运用同样的推理并在 (1.1.16) 式中令 $a=1$, 得到

$$\begin{aligned}u_1(s)&=\mathrm{E}[s^{\zeta_1},X_1=1]+\mathrm{E}[s^{\zeta_1},X_1=-1]\\&=ps+qs\mathrm{E}[s^{\zeta_2\circ\Sigma^1},X_1=-1]\\&=ps+qsu_2(s)=ps+qsu_1^2(s).\end{aligned}$$

因此, 由二次方程求根公式得

$$u_1(s)=\frac{1\pm\sqrt{1-4pqs^2}}{2qs}.$$

因为 P(ζ_1 为奇数) $=1$, 所以 $u_1(-s)=-u_1(s)$. 此外,

$$s\in(0,1)\implies\frac{1+\sqrt{1-4pqs^2}}{2qs}>\frac{1+\sqrt{1-4pq}}{2q}=\frac{p\vee q}{q}\geqslant 1.$$

[7] $a\vee b$ 表示最大值 $\max\{a,b\}, a,b\in\mathbb{R}$.

于是, 既然 $s \in (0,1)$ 意味着 $u_1(s) < 1$, 我们可以排除 "+" 解, 并因此完成 (1.1.10) 式的第二种推导. 事实上, 将它和 (1.1.16) 式相结合, 我们已经证明了当 $|s| < 1$ 时,

$$\mathrm{E}[s^{\zeta_a}] = \begin{cases} \left(\dfrac{1-\sqrt{1-4pqs^2}}{2qs}\right)^a, & \text{若 } a \in \mathbb{Z}^+; \\ \left(\dfrac{1-\sqrt{1-4pqs^2}}{2ps}\right)^{-a}, & \text{若 } -a \in \mathbb{Z}^+. \end{cases} \tag{1.1.17}$$

1.2 随机游动的常返性

在 §1.1.4, 我们研究了最近邻随机游动首次返回到 0 的时刻 ρ_0. 正如我们将在第二章和第三章看到的那样, 首次返回时刻对理解随机游动和相关过程的长时间行为起着关键的作用 (参看 §2.3.2). 事实上, 当随机游动回到 0 时, 它重新开始. 这样, 如果它以概率 1 返回, 那么随机游动的整个过程由许多时间区间组成, 每个时间区间是一条从 0 出发并在 0 结束的闭路径. 因为首次返回时刻标志着一个时间区间结束, 下一个同分布的时间区间重新开始, 所以它常被称作*常返时刻*. 并且如果 $\mathrm{P}(\rho_0 < \infty) = 1$, 那么称随机游动是*常返的*. 非常返的随机游动被称作*瞬时的*.

这一节, 我们将讨论最近邻随机游动的常返性质. 当然, 已经知道 (参看 (1.1.13)) $\mathbb{Z}$ 上最近邻随机游动是常返的当且仅当它是对称的. 因此, 我们的兴趣在于高维情形. 特别地, 为了确信常返性是相当微妙的, 我们将证明 $\mathbb{Z}^2$ 上的最近邻对称随机游动是常返的, 但 $\mathbb{Z}^3$ 上的最近邻对称随机游动则不然.

1.2.1. $\mathbb{Z}^d$上的随机游动: 为了描述 $\mathbb{Z}^d$ 上类似于 $\mathbb{Z}$ 上最近邻随机游动的性质, 首先把 $\boldsymbol{N}_1 \equiv \{-1, 1\}$ 看做 $\mathbb{Z}$ 上 0 的*最近邻集*. 容易明白, 为什么 $\mathbb{Z}^d$ 上原点的最近邻集由 $\mathbb{Z}^d$ 上的 $2d$ 个点组成, 其 $d-1$ 个坐标是 0, 余下的一个坐标是 $\boldsymbol{N}_1$ 中的点. 其次, 用 d 维独立同分布 $\boldsymbol{N}_d$ 值随机变量 $\boldsymbol{B}_1, \cdots, \boldsymbol{B}_n, \cdots$ [8] 取代 §1.1 $\boldsymbol{N}_1$ 值 Bernoulli 随机变量. 最

[8] $\boldsymbol{B}_n$ 的存在性可以看做是定理 6.3.2 的一个推论. 令 $\{U_n : n \in \mathbb{Z}^+\}$ 是一族服从 $[0,1)$ 上均匀分布的独立随机变量. 再令 $(\boldsymbol{k}_1, \cdots, \boldsymbol{k}_{2d})$ 是 $\boldsymbol{N}_d$ 中元素的一个排序, 令 $\beta_0 = 0$, $\beta_m = \sum_{\ell=1}^{m} \mathrm{P}(\boldsymbol{B}_1 = \boldsymbol{k}_\ell)$, $1 \leqslant m \leqslant 2d$, 定义 $F : [0,1) \to \boldsymbol{N}_d$ 使得 $F \upharpoonright [\beta_{m-1}, \beta_m) = \boldsymbol{k}_m$, 取 $\boldsymbol{B}_n = F(U_n)$.

后, $\mathbb{Z}^d$ 上最近邻随机游动是一族如下定义的随机变量 $\{\boldsymbol{X}_n : n \geqslant 0\}$,

$$\boldsymbol{X}_0 = \boldsymbol{0}, \quad \boldsymbol{X}_n = \sum_{m=1}^{n} \boldsymbol{B}_m, \quad n \geqslant 1.$$

等价地, $\{\boldsymbol{X}_n : n \geqslant 0\}$ 的具有随机过程特色的描述如下:

$$\begin{aligned} &\mathrm{P}(\boldsymbol{X}_0 = \boldsymbol{0}) = 1, \text{ 并且对 } n \geqslant 1 \text{ 和 } \boldsymbol{\epsilon} \in \boldsymbol{N}_d, \\ &\mathrm{P}(\boldsymbol{X}_n - \boldsymbol{X}_{n-1} = \boldsymbol{\epsilon} | \boldsymbol{X}_0, \cdots, \boldsymbol{X}_{n-1}) = p_{\boldsymbol{\epsilon}}, \end{aligned} \tag{1.2.1}$$

其中 $p_{\boldsymbol{\epsilon}} \equiv \mathrm{P}(\boldsymbol{B}_1 = \boldsymbol{\epsilon})$. 当 $\boldsymbol{B}_1$ 服从 $\boldsymbol{N}_d$ 上均匀分布时, 随机游动被称作**是对称的**.

和上面引进的概念和术语保持一致, 如果 $n \geqslant 1$, $\boldsymbol{X}_n = \boldsymbol{0}$, 并且对 $1 \leqslant m < n$, $\boldsymbol{X}_m \neq \boldsymbol{0}$, 那么定义首次返回原点的时刻 ρ_0 等于 n; 如果不存在这样的 $n \geqslant 1$, 那么取 $\rho_0 = \infty$. 另外, 将根据 $\mathrm{P}(\rho_0 < \infty)$ 等于 1 或者严格小于 1, 称随机游动是**常返的**或者**瞬时的**.

1.2.2. 一个初等的常返性判别法则: 给定 $n \geqslant 1$, 令 $\rho_{\boldsymbol{0}}^{(n)}$ 是第 n 次返回到 $\boldsymbol{0}$ 的时刻. 即$\rho_{\boldsymbol{0}}^{(1)} = \rho_{\boldsymbol{0}}$, 并且对 $n \geqslant 2$,

$$\rho_{\boldsymbol{0}}^{(n-1)} < \infty \implies \rho_{\boldsymbol{0}}^{(n)} = \inf\{m > \rho_{\boldsymbol{0}}^{(n-1)} : \boldsymbol{X}_m = \boldsymbol{0}\};$$

如果 $\rho_{\boldsymbol{0}}^{(n-1)} = \infty$, 那么 $\rho_{\boldsymbol{0}}^{(n)} = \infty$. 等价地, 如果定义 $g : (\boldsymbol{N}_d)^{\mathbb{Z}^+} \to \mathbb{Z}^+ \bigcup \{\infty\}$ 使得

$$g(\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_\ell, \cdots) > n, \quad \text{当且仅当} \sum_{\ell=1}^{m} \boldsymbol{\epsilon}_\ell \neq \boldsymbol{0} \text{ 对 } 1 \leqslant m \leqslant n \text{ 成立},$$

那么 $\rho_{\boldsymbol{0}} = g(\boldsymbol{B}_1, \cdots, \boldsymbol{B}_\ell, \cdots)$, 并且如果 $\rho_{\boldsymbol{0}}^{(n)} = m$, 那么 $\rho_{\boldsymbol{0}}^{(n+1)} = m + \rho_{\boldsymbol{0}} \circ \Sigma^m$, 其中 $\rho_{\boldsymbol{0}} \circ \Sigma^m$ 等于 $g(\boldsymbol{B}_{m+1}, \cdots, \boldsymbol{B}_{m+\ell}, \cdots)$. 特别地, 因为 $\{\rho_{\boldsymbol{0}}^{(n)} = m\}$ 仅依赖于 $(\boldsymbol{B}_1, \cdots, \boldsymbol{B}_m)$, 因此独立于 $\rho_{\boldsymbol{0}} \circ \Sigma^m$, 并且 $\rho_{\boldsymbol{0}} \circ \Sigma^m$和 $\rho_{\boldsymbol{0}}$ 具有相同的分布, 所以成立

$$\begin{aligned} \mathrm{P}(\rho_{\boldsymbol{0}}^{(n+1)} < \infty) &= \sum_{m=1}^{\infty} \mathrm{P}(\rho_{\boldsymbol{0}}^{(n)} = m, \rho_{\boldsymbol{0}} \circ \Sigma^m < \infty) \\ &= \mathrm{P}(\rho_{\boldsymbol{0}}^{(n)} < \infty)\mathrm{P}(\rho_{\boldsymbol{0}} < \infty). \end{aligned}$$

这样, 我们证明了: 对 $n \geqslant 1$,

$$\mathrm{P}(\rho_{\boldsymbol{0}}^{(n)} < \infty) = \mathrm{P}(\rho_{\boldsymbol{0}} < \infty)^n. \tag{1.2.2}$$

(1.2.2) 式的一个好处是, 它支持了前面给出的有关常返随机游动结构的解释. 也就是说, 如果随机游动一旦以概率 1 返回到 $\mathbf{0}$, 那么它将以概率 1 无穷多次返回到 $\mathbf{0}$. 该性质有许多应用. 例如, 如果 $\boldsymbol{B}_1$ 的第 i 个坐标的平均值不为 0, 那么 $\{\boldsymbol{X}_n : n \geqslant 0\}$ 一定是瞬时的. 为了说明这一点, 令 Y_n 表示 $\boldsymbol{B}_n$ 的第 i 个坐标, 注意到 $\{Y_n - Y_{n-1} : n \geqslant 1\}$ 是一列独立同分布 $\{-1, 0, 1\}$ 值随机变量, 均值为 $\mu \neq 0$. 但是, 由强大数律 (参看下面习题 1.3.4), 这意味着以概率 1 有 $\dfrac{Y_n}{n} \to \mu \neq 0$. 而这仅当 $|\boldsymbol{X}_n| \geqslant |Y_n| \to \infty$ 以概率 1 成立时才可能, 并且明显排除了, 即使以正概率, $\boldsymbol{X}_n = \mathbf{0}$ 无穷多次发生的可能性.

(1.2.2) 式的第二个好处如下. 定义

$$T_{\mathbf{0}} = \sum_{n=0}^{\infty} \mathbf{1}_{\{\mathbf{0}\}}(\boldsymbol{X}_n)$$

是 $\{\boldsymbol{X}_n : n \geqslant 0\}$ 停留在原点处的总时间. 因为 $\boldsymbol{X}_0 = \mathbf{0}$, 所以 $T_{\mathbf{0}} \geqslant 1$. 进而, 对 $n \geqslant 1$, $T_{\mathbf{0}} > n \Leftrightarrow \rho_{\mathbf{0}}^{(n)} < \infty$. 因此, 由 (1.2.2),

$$\begin{aligned}
\mathrm{E}[T_{\mathbf{0}}] &= \sum_{n=0}^{\infty} \mathrm{P}(T_{\mathbf{0}} > n) = 1 + \sum_{n=1}^{\infty} \mathrm{P}(\rho_{\mathbf{0}}^{(n)} < \infty) \\
&= 1 + \sum_{n=1}^{\infty} \mathrm{P}(\rho_{\mathbf{0}} < \infty)^n,
\end{aligned}$$

并由此得

$$\mathrm{E}[T_{\mathbf{0}}] = \frac{1}{1 - \mathrm{P}(\rho_{\mathbf{0}} < \infty)} = \frac{1}{\mathrm{P}(\rho_{\mathbf{0}} = \infty)}. \tag{1.2.3}$$

在将 (1.2.3) 式应用到常返性问题之前, 注意到 $T_{\mathbf{0}}$ 是一个具有下列特殊二分性的随机变量:

$$\begin{aligned}
&\mathrm{P}(T_{\mathbf{0}} < \infty) > 0 \implies \mathrm{E}[T_{\mathbf{0}}] < \infty, \\
&\mathrm{E}[T_{\mathbf{0}}] = \infty \implies \mathrm{P}(T_{\mathbf{0}} = \infty) = 1.
\end{aligned} \tag{1.2.4}$$

事实上, 如果 $\mathrm{P}(T_{\mathbf{0}} < \infty) > 0$, 那么 $\boldsymbol{X}_n$ 不可能以正概率无穷多次处于 $\mathbf{0}$. 因此根据 (1.1.13) 式 $\mathrm{P}(\rho_{\mathbf{0}} < \infty) < 1$, 进而根据 (1.2.3) 式, 可推出 $\mathrm{E}[T_{\mathbf{0}}] < \infty$. 另一方面, 如果 $\mathrm{E}[T_{\mathbf{0}}] = \infty$, 那么 (1.2.3) 式意味着 $\mathrm{P}(\rho_{\mathbf{0}} < \infty) = 1$, 并根据 (1.2.2) 式, 对所有 $n \geqslant 1$ 有 $\mathrm{P}(T_{\mathbf{0}} > n) = \mathrm{P}(\rho_{\mathbf{0}}^{(n)} < \infty) = 1$. 因此 $\mathrm{P}(T_{\mathbf{0}} = \infty) = 1$ (参看 (6.1.3)).

1.2.3. $\mathbb{Z}^2$ 上对称随机游动的常返性: 应用 (1.2.3) 式来确定常返性的最常用方法是利用公式

$$\mathrm{E}[T_{\mathbf{0}}] = \sum_{n=0}^{\infty} \mathrm{P}(\boldsymbol{X}_n = \mathbf{0}). \tag{1.2.5}$$

尽管 (1.2.5) 式的证明本质上是明显的 (参看定理 6.1.15):

$$\mathrm{E}[T_{\mathbf{0}}] = \mathrm{E}\left[\sum_{n=0}^{\infty} \mathbf{1}_{\{\mathbf{0}\}}(\boldsymbol{X}_n)\right] = \sum_{n=0}^{\infty} \mathrm{E}[\mathbf{1}_{\{\mathbf{0}\}}(\boldsymbol{X}_n)] = \sum_{n=0}^{\infty} \mathrm{P}(\boldsymbol{X}_n = \mathbf{0}),$$

但它与 (1.2.3) 式结合起来, 就变得很有用. 也就是说

$$\{\boldsymbol{X}_n : n \geqslant 0\} \text{ 常返当且仅当 } \sum_{n=0}^{\infty} \mathrm{P}(\boldsymbol{X}_n = \mathbf{0}) = \infty, \tag{1.2.6}$$

并且因为 $\mathrm{P}(\boldsymbol{X}_n = \mathbf{0})$ 比涉及多个时刻路径性质的量更容易估计, 这提供了很有价值的信息.

为了将 (1.2.6) 式应用到对称随机游动上, 注意到下列事实是非常重要的: *当随机游动对称时, 在任何偶数时刻, 随机游动处于 $\mathbf{0}$ 点的可能性最大.* 为了验证这一点, 注意到如果 $\boldsymbol{k} \in \mathbb{Z}^d$, 那么

$$\begin{aligned}
\mathrm{P}(\boldsymbol{X}_{2n} = \boldsymbol{k}) &= \sum_{\boldsymbol{\ell} \in \mathbb{Z}^d} \mathrm{P}(\boldsymbol{X}_n = \boldsymbol{\ell}, \boldsymbol{X}_{2n} - \boldsymbol{X}_n = \boldsymbol{k} - \boldsymbol{\ell}) \\
&= \sum_{\boldsymbol{\ell} \in \mathbb{Z}^d} \mathrm{P}(\boldsymbol{X}_n = \boldsymbol{\ell}) \mathrm{P}(\boldsymbol{X}_{2n} - \boldsymbol{X}_n = \boldsymbol{k} - \boldsymbol{\ell}) \\
&= \sum_{\boldsymbol{\ell} \in \mathbb{Z}^d} \mathrm{P}(\boldsymbol{X}_n = \boldsymbol{\ell}) \mathrm{P}(\boldsymbol{X}_n = \boldsymbol{k} - \boldsymbol{\ell}) \\
&\leqslant \left(\sum_{\boldsymbol{\ell} \in \mathbb{Z}^d} \mathrm{P}(\boldsymbol{X}_n = \boldsymbol{\ell})^2\right)^{\frac{1}{2}} \left(\sum_{\boldsymbol{\ell} \in \mathbb{Z}^d} \mathrm{P}(\boldsymbol{X}_n = \boldsymbol{k} - \boldsymbol{\ell})^2\right)^{\frac{1}{2}} \\
&= \sum_{\boldsymbol{\ell} \in \mathbb{Z}^d} \mathrm{P}(\boldsymbol{X}_n = \boldsymbol{\ell})^2,
\end{aligned}$$

其中最后一行我们应用了 Schwarz 不等式 (参看下面习题 1.3.1). 至此, 我们还没有利用对称性. 然而, 如果随机游动是对称的, 那么 $\mathrm{P}(\boldsymbol{X}_n = \boldsymbol{\ell}) = \mathrm{P}(\boldsymbol{X}_n = -\boldsymbol{\ell})$, 并由此, 前面最后一行可以继续写成

$$\begin{aligned}
&\sum_{\boldsymbol{\ell} \in \mathbb{Z}^d} \mathrm{P}(\boldsymbol{X}_n = \boldsymbol{\ell}) \mathrm{P}(\boldsymbol{X}_n = -\boldsymbol{\ell}) \\
&= \sum_{\boldsymbol{\ell} \in \mathbb{Z}^d} \mathrm{P}(\boldsymbol{X}_n = \boldsymbol{\ell}) \mathrm{P}(\boldsymbol{X}_{2n} - \boldsymbol{X}_n = -\boldsymbol{\ell}) = \mathrm{P}(\boldsymbol{X}_{2n} = \mathbf{0}).
\end{aligned}$$

这样,

$$\{\boldsymbol{X}_n : n \geqslant 0\} \text{ 对称} \implies \mathrm{P}(\boldsymbol{X}_{2n} = \boldsymbol{0}) = \max_{\boldsymbol{k} \in \mathbb{Z}^d} \mathrm{P}(\boldsymbol{X}_{2n} = \boldsymbol{k}). \quad (1.2.7)$$

为了理解如何应用上述想法, 我们首先给出 $\mathbb{Z}$ 上最近邻对称随机游动常返性的第二种推导. 为此, 注意到因为 $\mathrm{P}(|X_n| \leqslant n) = 1$, 所以 (1.2.7) 式推出

$$1 = \sum_{\ell=-2n}^{2n} \mathrm{P}(X_{2n} = \ell) \leqslant (4n+1)\mathrm{P}(X_{2n} = 0),$$

因此由调和级数的发散性得 $\sum\limits_{n=0}^{\infty} \mathrm{P}(X_{2n} = 0) = \infty$.

$\mathbb{Z}^2$ 上最近邻对称随机游动的分析还需要用到其他的内容. 即在 d 维情况下, 类似于前面的推理可以得到 $\mathrm{P}(\boldsymbol{X}_{2n} = \boldsymbol{0}) \geqslant (4n+1)^{-d}$. 除 $d = 1$ 外, 该式不能给出确切的结论. 为了做得更好些, 我们需要利用事实

$$\{\boldsymbol{X}_n : n \geqslant 0\} \text{ 对称} \implies \mathrm{E}[|\boldsymbol{X}_n|^2] = n. \quad (1.2.8)$$

为了证明 (1.2.8) 式, 我们指出 $\boldsymbol{B}_n$ 的每一个坐标都是均值为 0、方差为 $\frac{1}{d}$ 的随机变量. 因此, 由于 $\boldsymbol{B}_n$ 相互独立, $\boldsymbol{X}_n$ 的每个坐标的二阶矩为 $\frac{n}{d}$.

有了 (1.2.8) 式, Markov 不等式 (6.1.12) 表明

$$\mathrm{P}(|\boldsymbol{X}_{2n}| \geqslant \sqrt{2n}) \leqslant \frac{1}{4n}\mathrm{E}[|\boldsymbol{X}_{2n}|^2] = \frac{1}{2},$$

它可以帮助改进前面的讨论, 得到[9]

$$\begin{aligned}\frac{1}{2} \leqslant \mathrm{P}(|\boldsymbol{X}_{2n}| < \sqrt{2n}) &= \sum_{|\boldsymbol{\ell}| < \sqrt{2n}} \mathrm{P}(\boldsymbol{X}_{2n} = \boldsymbol{\ell}) \\ &\leqslant (\sqrt{8n}+1)^d \mathrm{P}(\boldsymbol{X}_{2n} = \boldsymbol{0}) \leqslant 2^{d-1}((8n)^{\frac{d}{2}}+1)\mathrm{P}(\boldsymbol{X}_{2n} = \boldsymbol{0}).\end{aligned}$$

这就是说, 我们已经对 $\mathbb{Z}^d$ 上最近邻对称随机游动证明了

$$\mathrm{P}(\boldsymbol{X}_{2n} = \boldsymbol{0}) \geqslant 2^{-d}((8n)^{\frac{d}{2}}+1)^{-1}. \quad (1.2.9)$$

[9] 对任何 $a, b \in [0, \infty)$ 和 $p \in [1, \infty)$, $(a+b)^p \leqslant 2^{p-1}(a^p + b^p)$. 这可以被看做是 Jensen 不等式的应用 (参看习题 5.6.2). 简单地说, $x \in [0, \infty) \mapsto x^p$ 是一个凸函数.

特别地, 当 $d=2$ 时, 上式表明$\mathbb{Z}^2$ 上最近邻对称随机游动是常返的.

1.2.4. $\mathbb{Z}^3$ 上的瞬时性: 尽管由 (1.2.9) 式可以证明 $\mathbb{Z}^2$ 上最近邻对称随机游动的常返性, 但它没有解决 $d \geqslant 3$ 时 $\mathbb{Z}^d$ 上随机游动的瞬时性问题. 因此, 为了处理 $d \geqslant 3$ 情形, 我们需要分析 (1.2.9) 式的估计究竟有多好. 实际上, 只要证明存在同样形式的上界即可.

为了获得与 (1.2.9) 式中下界相对应的上界, 我们首先考察 $d=1$ 的情形. 为此, 给定 $0 \leqslant \ell \leqslant n$, 注意到

$$\begin{aligned}\frac{\mathrm{P}(X_{2n}=2\ell)}{\mathrm{P}(X_{2n}=0)} &= \frac{(n!)^2}{(n+\ell)!(n-\ell)!} = \frac{n(n-1)\cdots(n-\ell+1)}{(n+\ell)(n+\ell-1)\cdots(n+1)} \\ &= \prod_{k=0}^{\ell-1}\left(1-\frac{\ell}{n+\ell-k}\right) \geqslant \left(1-\frac{\ell}{n+1}\right)^{\ell}.\end{aligned}$$

回顾

$$\log(1-x) = -\sum_{m=1}^{\infty}\frac{x^m}{m}, \quad |x|<1, \tag{1.2.10}$$

并由此知, 对 $0 \leqslant x \leqslant \frac{1}{2}$ 有 $\log(1-x) \geqslant -\frac{3x}{2}$. 因此, 上述表明只要 $0 \leqslant \ell \leqslant \frac{n+1}{2}$, 就有

$$\frac{\mathrm{P}(X_{2n}=2\ell)}{\mathrm{P}(X_{2n}=0)} \geqslant e^{-\ell\log\left(1-\frac{\ell}{n+1}\right)} \geqslant e^{-\frac{3\ell^2}{2(n+1)}}.$$

因为 $\mathrm{P}(X_{2n}=-2\ell)=\mathrm{P}(X_{2n}=2\ell)$, 所以

$$\mathrm{P}(X_{2n}=0) \leqslant e^{\frac{3}{2}}\mathrm{P}(X_{2n}=2\ell), \quad |\ell| \leqslant \sqrt{n}.$$

但是, 由 $\sum_{\ell}\mathrm{P}(X_{2n}=2\ell)=1$ 可推出 $(2\sqrt{n}-1)\mathrm{P}(X_{2n}=0) \leqslant e^{\frac{3}{2}}$. 因此当 $\{X_n : n \geqslant 0\}$ 是 $\mathbb{Z}$ 上最近邻对称随机游动时,

$$\mathrm{P}(X_{2n}=0) \leqslant e^{\frac{3}{2}}(2\sqrt{n}-1)^{-1}, \quad n \geqslant 1. \tag{1.2.11}$$

如果最近邻对称随机游动的坐标相互独立 (大多数肯定不是), 那么 (1.2.11) 式给出我们所要的上界. 这样, 有必要看看在多大程度上我们能把 $\mathbb{Z}^d$ 上最近邻对称随机游动和 $\mathbb{Z}$ 上 d 个相互独立的最近邻对称随机游动 $\{X_{i,n} : n \geqslant 0\}, 1 \leqslant i \leqslant d$, 联系起来. 为此, 参看 (1.2.1) 式, 其中 $p_\epsilon \equiv \frac{1}{2d}$, 并分两步选择 $\boldsymbol{X}_n - \boldsymbol{X}_{n-1}$: 首先选择非零的坐

标, 接着选择它是否等于 $+1$ 或者 -1. 记住这一点, 令 $\{I_n : n \geqslant 0\}$ 是一列 $\{1, 2, \cdots, d\}$ 值、相互独立均匀分布的随机变量, 并且独立于 $\{X_{i,n} : 1 \leqslant i \leqslant d, n \geqslant 0\}$. 对 $1 \leqslant i \leqslant d$, 令 $N_{i,0} = 0$, 并且当 $n \geqslant 1$ 时, 令 $N_{i,n} = \sum\limits_{m=1}^{n} \mathbf{1}_{\{i\}}(I_m)$. 考虑如下定义的序列 $\{\boldsymbol{Y}_n : n \geqslant 0\}$

$$\boldsymbol{Y}_n = (X_{1,N_{1,n}}, \cdots, X_{d,N_{d,n}}). \tag{1.2.12}$$

不难验证 $\{\boldsymbol{Y}_n : n \geqslant 0\}$ 满足 $\mathbb{Z}^d$ 上最近邻对称随机游动的条件 (1.2.1), 因此和 $\{\boldsymbol{X}_n : n \geqslant 0\}$ 具有相同的分布. 特别地, 根据 (1.2.11),

$$\begin{aligned}
&\mathrm{P}(\boldsymbol{X}_{2n} = \mathbf{0}) \\
&= \sum_{\boldsymbol{m} \in \mathbb{N}^d} \mathrm{P}(X_{i,2m_i} = 0, N_{i,2n} = 2m_i, 1 \leqslant i \leqslant d) \\
&= \sum_{\substack{\boldsymbol{m} \in \mathbb{N}^d \\ m_1 \wedge \cdots \wedge m_d \geqslant \frac{n}{d}}} \left(\prod_{i=1}^{d} \mathrm{P}(X_{i,2m_i} = 0) \right) \mathrm{P}(N_{i,2n} = 2m_i, 1 \leqslant i \leqslant d) + \\
&\quad \sum_{\substack{\boldsymbol{m} \in \mathbb{N}^d \\ m_1 \wedge \cdots \wedge m_d < \frac{n}{d}}} \left(\prod_{i=1}^{d} \mathrm{P}(X_{i,2m_i} = 0) \right) \mathrm{P}(N_{i,2n} = 2m_i, 1 \leqslant i \leqslant d) \\
&\leqslant e^{\frac{3d}{2}} \left(2\sqrt{\frac{n}{d}} - 1 \right)^{-d} + \mathrm{P}\left(N_{i,2n} \leqslant \frac{n}{d}, \text{对某个 } 1 \leqslant i \leqslant d \right).
\end{aligned}$$

这样, 一旦证明了存在常数 $B(d) < \infty$ 使得

$$\mathrm{P}\left(N_{i,2n} \leqslant \frac{n}{d}, \text{对某个 } 1 \leqslant i \leqslant d \right) \leqslant B(d) n^{-\frac{d}{2}}, \quad n \geqslant 1, \tag{1.2.13}$$

那么就可以证明存在常数 $A(d) < \infty$ 使得

$$\mathrm{P}(\boldsymbol{X}_{2n} = \mathbf{0}) \leqslant A(d) n^{-\frac{d}{2}}, \quad n \geqslant 1. \tag{1.2.14}$$

特别地, 将得到

$$d \geqslant 3 \implies \sum_{n=0}^{\infty} \mathrm{P}(\boldsymbol{X}_{2n} = \mathbf{0}) < \infty.$$

因此当 $d \geqslant 3$ 时, $\mathbb{Z}^d$ 上最近邻对称随机游动是瞬时的.

为了证明 (1.2.13) 式, 首先注意到

$$\mathrm{P}\left(N_{i,2n} \leqslant \frac{n}{d}, \text{对某个 } 1 \leqslant i \leqslant d \right) \leqslant d\mathrm{P}\left(N_{1,2n} \leqslant \frac{n}{d} \right).$$

下面令 $N_{1,n} = \sum_1^n Z_m$, 其中 $Z_m = \mathbf{1}_{\{1\}}(I_m)$, 并注意到 $\{Z_m : m \geqslant 1\}$ 是一列 $\{0,1\}$ 值 Bernoulli 随机变量, $\mathrm{P}(Z_m = 1) = p \equiv \frac{1}{d}$. 特别地, 对任意 $\lambda \in \mathbb{R}$,

$$\mathrm{E}\left[\exp\left(\lambda \sum_1^n Z_m\right)\right] = (pe^{\lambda} + q)^n,$$

因此

$$\mathrm{E}\left[\exp\left(\lambda\left(np - \sum_1^n Z_m\right)\right)\right] = e^{n\psi(\lambda)},$$

$$\text{其中 } \psi(\lambda) \equiv \log\left(pe^{-\lambda q} + qe^{\lambda p}\right).$$

因为 $\psi(0) = \psi'(0) = 0$, 并且

$$\psi''(\lambda) = \frac{pqe^{\lambda(p-q)}}{(qe^{\lambda p} + pe^{-\lambda q})^2} = \frac{pq}{(qx_\lambda + px_\lambda^{-1})^2} \leqslant \frac{1}{4},$$

其中 $x_\lambda \equiv e^{\frac{1}{2}\lambda(p+q)}$, 所以由 Taylor 公式得

$$\mathrm{E}\left[\exp\left(\lambda\left(np - \sum_1^n Z_m\right)\right)\right] \leqslant e^{\frac{n\lambda^2}{8}}, \quad \lambda \in \mathbb{R}. \tag{1.2.15}$$

从 (1.2.15) 式出发, 有多种方法可以推出 (1.2.13) 式. 例如, 对任何 $\lambda > 0$ 和 $R > 0$, 由 Markov 不等式 (6.1.12) 和 (1.2.15) 式可推出

$$\begin{aligned}\mathrm{P}\left(\sum_1^n Z_m \leqslant np - nR\right) &= \mathrm{P}\left(\exp\left(\lambda\left(np - \sum_1^n Z_m\right)\right) \geqslant e^{n\lambda R}\right) \\ &\leqslant e^{-n\lambda R + \frac{n\lambda^2}{8}},\end{aligned}$$

当 $\lambda = 4nR$ 时, 上式给出了

$$\mathrm{P}\left(\sum_1^n Z_m \leqslant np - nR\right) \leqslant e^{-2nR^2}. \tag{1.2.16}$$

回到以前使用过的记号, 并注意到有关 (1.2.13) 式的说明, 从 (1.2.16) 式可以推出

$$\mathrm{P}\left(N_{i,2n} \leqslant \frac{n}{d}, \text{ 对某个 } 1 \leqslant i \leqslant d\right) \leqslant de^{-\frac{2n^2}{d^2}},$$

它明显比 (1.2.13) 式要好很多.

我们在这一小节里使用的论证是一种极为有用的被称作耦合的方法. 粗略地说, 耦合方法要求把人们希望进一步了解的随机变量 (这里是 $\mathbf{X}_{2n}$) 写作某些已经了解得相当多的随机变量 (这里是 $\{(X_{i,m}, N_{i,m}): 1 \leqslant i \leqslant d, m \geqslant 0\}$) 的函数. 当然, 方法的好坏反映着使用者的功力: 存在许多方法把一个随机变量耦合成其他随机变量, 但大多数并没有用.

1.3 习题

1.3.1. 存在多种形式的 *Schwarz* 不等式, 其中最初等的是, 对任何 $\{a_n : n \in \mathbb{Z}\} \subseteq \mathbb{R}$ 并且 $\{b_n : n \in \mathbb{Z}\} \subseteq \mathbb{R}$,

$$\sum_{n\in\mathbb{Z}} |a_n b_n| \leqslant \sqrt{\sum_{n\in\mathbb{Z}} a_n^2}\sqrt{\sum_{n\in\mathbb{Z}} b_n^2}.$$

进而, 当右边有限时,

$$\left|\sum_{n\in\mathbb{Z}} a_n b_n\right| = \sqrt{\sum_{n\in\mathbb{Z}} a_n^2}\sqrt{\sum_{n\in\mathbb{Z}} b_n^2}$$

当且仅当存在一个 $\alpha \in \mathbb{R}$ 有 $b_n = \alpha a_n, n \in \mathbb{Z}$ 或者 $a_n = \alpha b_n, n \in \mathbb{Z}$. 下面是这些结论的证明提纲.

(a) 首先说明只要考虑下列情况即可: 除有限个外, 对所有的 n, $a_n = 0 = b_n$.

(b) 给定实二次多项式 $P(x) = Ax^2 + 2Bx + C$, 利用二次求根公式得, P 处处大于或等于 0 当且仅当 $A + C \geqslant 0$ 并且 $B^2 \leqslant AC$. 类似地, 证明 P 处处大于 0 当且仅当 $A > 0$ 并且 $B^2 < AC$.

(c) 假设除有限个外, 对所有的 n, $a_n = 0 = b_n$. 令 $P(x) = \sum_n (a_n x + b_n)^2$, 应用 (b) 得到所要的结论. 最后, 利用 (a) 去掉对 a_n 和 b_n 的限制.

1.3.2. 令 $\{Y_n : n \geqslant 1\}$ 是一列相互独立同分布的随机变量, 满足 $\mathrm{E}[|Y_1|] < \infty$. 对 $n \geqslant 1$, 令 $X_n = \sum_{m=1}^{n} Y_m$. 弱大数律表明

$$\mathrm{P}\left(\left|\frac{X_n}{n} - \mathrm{E}[Y_1]\right| \geqslant \epsilon\right) \to 0, \quad 对所有 \epsilon > 0.$$

事实上,

$$\lim_{n\to\infty} \mathrm{E}\left[\left|\frac{X_n}{n} - \mathrm{E}[Y_1]\right|\right] = 0, \tag{1.3.3}$$

从上式出发, 作为 Markov 不等式的应用, 即得弱大数律成立. 下面是 (1.3.3) 式的证明步骤.

(a) 首先简化成 $\mathrm{E}[Y_1]=0$ 的情形. 然后假设 $\mathrm{E}[Y_1^2]<\infty$, 并证明

$$\mathrm{E}\left[\left|\frac{X_n}{n}\right|\right]^2 \leqslant \mathrm{E}\left[\left|\frac{X_n}{n}\right|^2\right] = \frac{\mathrm{E}[Y_1^2]}{n}.$$

从而, 当 Y_1 存在有限二阶矩时, 结论得以证明.

(b) 给定 $R>0$, 令 $Y_n^{(R)}=Y_n\mathbf{1}_{[0,R)}(|Y_n|)-\mathrm{E}[Y_n,|Y_n|<R]$, $X_n^{(R)}=\sum\limits_{m=1}^{n}Y_m^{(R)}$. 注意到, 对任何 $R>0$,

$$\begin{aligned}\mathrm{E}\left[\left|\frac{X_n}{n}\right|\right] &\leqslant \mathrm{E}\left[\left|\frac{X_n^{(R)}}{n}\right|\right]+\mathrm{E}\left[\left|\frac{X_n-X_n^{(R)}}{n}\right|\right]\\ &\leqslant \sqrt{\mathrm{E}\left[\left(\frac{X_n^{(R)}}{n}\right)^2\right]}+2\mathrm{E}[Y_1,|Y_1|\geqslant R]\\ &\leqslant \frac{R}{n^{1/2}}+2\mathrm{E}[Y_1,|Y_1|\geqslant R].\end{aligned}$$

应用上面不等式, 结合单调收敛定理, 完成 (1.3.3) 式的证明.

1.3.4. 参看习题 1.3.2. 强大数律表明弱大数律能被改进为: $\dfrac{X_n}{n}\to\mathrm{E}[Y_1]$ 以概率 1 成立. 如果仅仅假设 $\mathrm{E}[|Y_1|]<\infty$, 强大数律的证明需要一些技巧. 然而, 如果假设 $\mathrm{E}[Y_1^4]<\infty$, 那么可以基于和弱大数律同样的讨论来证明.

令 $\{Y_n\}_1^\infty$ 是一列独立同分布的随机变量, 满足 $M=\sup\limits_n\mathrm{E}[|Y_n|^4]<\infty$. 证明以概率 1 成立 $\lim\limits_{n\to\infty}\dfrac{1}{n}\sum\limits_{m=1}^{n}(Y_m-\mathrm{E}[Y_m])=0$. 注意, 我们并没有假设它们是同分布的, 但是如果添加这一假设, 我们得到 $\lim\limits_{n\to\infty}\dfrac{1}{n}\sum\limits_{m=1}^{n}Y_m=\mathrm{E}[Y_1]$ 以概率 1 成立.

以下是证明提纲.

(a) 首先简化成对所有 $n\in\mathbb{Z}^+$, $\mathrm{E}[Y_n]=0$ 的情形.

(b) 接着进行展开

$$\mathrm{E}\left[\left(\sum_1^n Y_k\right)^4\right]=\sum_{k_1,\cdots,k_4=1}^{n}\mathrm{E}[Y_{k_1}\cdots Y_{k_4}],$$

并注意到不为 0 的项是那些每个指标至少等于一个其他指标的项. 从而得出

$$\mathrm{E}\left[\left(\sum_1^n Y_k\right)^4\right] = \sum_{k=1}^n \mathrm{E}[Y_k^4] + 6 \sum_{1\leqslant k<\ell\leqslant n} \mathrm{E}[Y_k^2]\mathrm{E}[Y_\ell^2].$$

因此, 由 $\mathrm{E}[Y_k^2]^2 \leqslant \mathrm{E}[Y_k^4]$, 得到

$$\mathrm{E}\left[\left(\sum_1^n Y_k\right)^4\right] \leqslant 3Mn^2. \tag{$*$}$$

(**c**) 从 $(*)$ 出发, 证明对所有 $\epsilon > 0$,

$$\mathrm{P}\left(\left|\frac{\sum_1^n Y_k}{n}\right| \geqslant \epsilon\right) \leqslant \frac{1}{\epsilon^4}\mathrm{E}\left[\left|\frac{\sum_1^n Y_k}{n}\right|^4\right] \leqslant \frac{3M}{\epsilon^4 n^2} \to 0.$$

这正是相互独立具有有界四阶矩的随机变量的弱大数律. 当然, 这里四阶矩的使用有点可笑, 因为仅使用二阶矩更容易些.

(**d**) 再次从 $(*)$ 出发并使用 (6.1.4) 式, 证明对所有 $\epsilon > 0$, 当 $m \to \infty$ 时,

$$\begin{aligned}\mathrm{P}\left(\sup_{n>m}\left|\frac{\sum_1^n Y_k}{n}\right| \geqslant \epsilon\right) &\leqslant \sum_{n=m+1}^{\infty} \mathrm{P}\left(\left|\sum_1^n Y_k\right| \geqslant n\epsilon\right)\\ &\leqslant \frac{4M}{\epsilon^4}\sum_{n=m+1}^{\infty}\frac{1}{n^2} \leqslant \frac{4M}{\epsilon^4 m} \to 0.\end{aligned}$$

(**e**) 运用收敛性定义和 (6.1.4) 式, 证明

$$\begin{aligned}\mathrm{P}\left(\frac{\sum_1^n Y_k}{n} \nrightarrow 0\right) &= \mathrm{P}\left(\bigcup_{N=1}^{\infty}\bigcap_{m=1}^{\infty}\bigcup_{n>m}\left|\frac{\sum_1^n Y_k}{n}\right| \geqslant \frac{1}{N}\right)\\ &\leqslant \sum_{N=1}^{\infty}\mathrm{P}\left(\bigcap_{m=1}^{\infty}\bigcup_{n>m}\left|\frac{\sum_1^n Y_k}{n}\right| \geqslant \frac{1}{N}\right).\end{aligned}$$

最后, 应用 (6.1.3) 式的第二行和上述 (**d**), 证明对每个 $N \in \mathbb{Z}^+$,

$$\mathrm{P}\left(\bigcap_{m=1}^{\infty}\bigcup_{n>m}\left|\frac{\sum_1^n Y_k}{n}\right| \geqslant \frac{1}{N}\right) = \lim_{m\to\infty}\mathrm{P}\left(\bigcup_{n>m}\left|\frac{\sum_1^n Y_k}{n}\right| \geqslant \frac{1}{N}\right) = 0.$$

因此, $\frac{1}{n}\sum_1^n Y_k \to 0$ 以概率 1 成立, 这正是独立具有有界四阶矩的随机变量的强大数律.

1.3.5. 熟悉中心极限定理的 DeMoivre 证明的读者可能已经注意到, (1.2.11) 式的估计是 *Stirling* 公式的一个很差的结果. Stirling 公式是

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n, \quad n \to \infty, \tag{1.3.6}$$

其中 $\sim$ 意味着两边数的比趋于 1. 给定 (1.3.6) 式, 证明

$$\mathrm{P}(X_{2n} = 0) \sim \sqrt{\frac{2}{\pi n}}.$$

然后, 基于以下提纲给出 (1.3.6) 式的证明.

(a) 令 $\tau_1, \cdots, \tau_n$ 是独立单位指数随机变量[10], 证明对任何 $0 < R \leqslant \sqrt{n}$,

$$\begin{aligned}1 - \frac{1}{R^2} &\leqslant \mathrm{P}\left(-R \leqslant \frac{\tau_1 + \cdots + \tau_n - n}{\sqrt{n}} \leqslant R\right) \\ &= \frac{1}{(n-1)!}\int_{-\sqrt{n}R+n}^{\sqrt{n}R+n} t^{n-1}e^{-t}dt.\end{aligned}$$

(b) 作变量替换, 并进行初等运算得到

$$\begin{aligned}\int_{-\sqrt{n}R+n}^{\sqrt{n}R+n} t^{n-1}e^{-t}dt &= \sqrt{n}e^{-n}\int_{-R}^{R}(n+\sqrt{n}\sigma)^{n-1}e^{-\sqrt{n}\sigma}d\sigma \\ &= n^{n-\frac{1}{2}}e^{-n}\int_{-R}^{R}\left(1+\frac{\sigma}{\sqrt{n}}\right)^{n-1}e^{-\sqrt{n}\sigma}d\sigma \\ &= n^{n-\frac{1}{2}}e^{-n}\int_{-R}^{R}\exp\left(-\frac{\sigma^2}{2}+E_n(\sigma)\right)d\sigma,\end{aligned}$$

其中

$$E_n(\sigma) \equiv (n-1)\log\left(1+\frac{\sigma}{\sqrt{n}}\right) - \sqrt{n}\sigma + \frac{\sigma^2}{2}.$$

(c) 作为 $\log(1+x)$ 的 Taylor 级数的应用 (参看 (1.2.10)), 证明当 $n \to \infty$ 时, $E_n(\sigma) \to 0$ 对 $|\sigma| \leqslant R$ 一致成立. 把它和 (a)、(b) 中的结果结合起来得到

$$\limsup_{n\to\infty} \frac{n^{n+\frac{1}{2}}e^{-n}}{n!}\int_{-R}^{R}\exp\left(-\frac{\sigma^2}{2}\right)d\sigma \leqslant 1$$

[10] 单位指数随机变量 τ 满足 $\mathrm{P}(\tau > t) = e^{-t\vee 0}$.

和

$$\limsup_{n\to\infty}\frac{n^{n+\frac{1}{2}}e^{-n}}{n!}\int_{-R}^{R}\exp\left(-\frac{\sigma^2}{2}\right)d\sigma\geqslant 1-\frac{1}{R^2}.$$

因为 $\int_{-\infty}^{\infty}\exp\left(-\frac{\sigma^2}{2}\right)d\sigma=\sqrt{2\pi}$, 那么令 $R\nearrow\infty$, (1.3.6) 式明显成立.

1.3.7. §1.2.3 的讨论是相当稳健的. 事实上, 令 $\{\boldsymbol{X}_n : n\geqslant 0\}$ 是 $\mathbb{Z}^2$ 上的任意对称随机游动, 其跳跃具有有限二阶矩. 即 $\boldsymbol{X}_0=\mathbf{0}$, $\{\boldsymbol{X}_n-\boldsymbol{X}_{n-1} : n\geqslant 1\}$ 是相互独立同分布、对称 ($\boldsymbol{X}_1$ 和 $-\boldsymbol{X}_1$ 同分布)、$\mathbb{Z}^2$ 值的随机变量. 具有有限二阶矩. 证明 $\{\boldsymbol{X}_n : n\geqslant 0\}$ 是常返的: P (存在 $n\geqslant 1, \boldsymbol{X}_n=\mathbf{0})=1$.

1.3.8. 令 $\{\boldsymbol{X}_n : n\geqslant 0\}$ 是 $\mathbb{Z}^d$ 上的随机游动: $\boldsymbol{X}_0=\mathbf{0}$, $\{\boldsymbol{X}_n-\boldsymbol{X}_{n-1} : n\geqslant 1\}$ 是相互独立同分布、$\mathbb{Z}^d$ 值的随机变量. 进而, 对 $1\leqslant i\leqslant d$, 令 $(\boldsymbol{X}_n)_i$ 是 $\boldsymbol{X}_n$ 的第 i 个坐标, 并且假设

$$\min_{1\leqslant i\leqslant d}\mathrm{P}((\boldsymbol{X}_1)_i\neq 0)>0,\quad \text{但是 P (存在 } i\neq j, (\boldsymbol{X}_1)_i(\boldsymbol{X}_1)_j\neq 0)=0.$$

如果对某个 $C<\infty$ 和满足 $\sum\limits_1^d\alpha_i>1$ 的 $(\alpha,\cdots,\alpha_d)\in[0,\infty)^d$, $\mathrm{P}((\boldsymbol{X}_n)_i=0)\leqslant Cn^{-\alpha_i}, n\geqslant 1$, 证明 $\{\boldsymbol{X}_n : n\geqslant 0\}$ 是瞬时的: P (存在 $n\geqslant 1, \boldsymbol{X}_n=\mathbf{0})<1$.

1.3.9. 正如前面一样, 令 $\{\boldsymbol{X}_n : n\geqslant 0\}$ 是 $\mathbb{Z}^d$ 上的随机游动. 给定 $\boldsymbol{k}\in\mathbb{Z}^d$, 令

$$T_{\boldsymbol{k}}=\sum_{n=0}^{\infty}\mathbf{1}_{\{\boldsymbol{k}\}}(\boldsymbol{X}_n),\quad \zeta_{\boldsymbol{k}}=\inf\{n\geqslant 0 : \boldsymbol{X}_n=\boldsymbol{k}\}.$$

证明

$$\mathrm{E}[T_{\boldsymbol{k}}]=\mathrm{P}(\zeta_{\boldsymbol{k}}<\infty)\mathrm{E}[T_0]=\frac{\mathrm{P}(\zeta_{\boldsymbol{k}}<\infty)}{\mathrm{P}(\rho_{\mathbf{0}}=\infty)},\tag{1.3.10}$$

其中 $\rho_{\mathbf{0}}=\inf\{n\geqslant 1 : \boldsymbol{X}_n=\mathbf{0}\}$ 是首次返回到 $\mathbf{0}$ 的时刻. 特别地, 如果 $\{\boldsymbol{X}_n : n\geqslant 0\}$ 在上题描述的意义下是瞬时的, 证明对所有 $r\in(0,\infty)$,

$$\mathrm{E}\left[\sum_{m=0}^{\infty}\mathbf{1}_{B(r)}(\boldsymbol{X}_n)\right]<\infty,$$

其中 $B(r)=\{\boldsymbol{k} : |\boldsymbol{k}|\leqslant r\}$; 由此得出 $|\boldsymbol{X}_n|\to\infty$ 以概率 1 成立. 另一方面, 如果 $\{\boldsymbol{X}_n : n\geqslant 0\}$ 是常返的, 证明 $\boldsymbol{X}_n=\mathbf{0}$ 以概率 1 无穷多次

发生. 因此, $\{\boldsymbol{X}_n : n \geqslant 0\}$ 是常返的并且 $\boldsymbol{X}_n = \boldsymbol{0}$ 以概率 1 无穷多次发生, 或者是瞬时的并且 $|\boldsymbol{X}_n| \to \infty$ 以概率 1 成立.

1.3.11. 在上题中, 令 $d = 1$, $X_0 = 0$, 并且 $\{X_n - X_{n-1} : n \geqslant 1\}$ 是独立同分布的随机变量, $0 < \mathrm{E}[|X_1|] < \infty$ 和 $\mathrm{E}[X_1] = 0$. 将 §1.2.1 的讨论作少许变化, 我们来证明该随机游动是常返的, 但是以概率 1 有

$$\limsup_{n\to\infty} X_n = \infty, \quad \liminf_{n\to\infty} X_n = -\infty.$$

(a) 首先说明只要证明 $\sup_n X_n = \infty$ 和 $\inf_n X_n = -\infty$ 即可. 然后, 利用弱大数律 (参看习题 1.3.2) 证明

$$\lim_{n\to\infty} \max_{1\leqslant m\leqslant n} \frac{\mathrm{E}[|X_m|]}{n} = 0.$$

(b) 对 $n \geqslant 1$, 令 $T_k^{(n)} = \sum_{m=0}^{n-1} \mathbf{1}_{\{k\}}(X_m)$, 证明对所有 $k \in \mathbb{Z}$, $\mathrm{E}\left[T_k^{(n)}\right] \leqslant \mathrm{E}\left[T_0^{(n)}\right]$, 并利用这个不等式得

$$(4\mu(n)+1)\mathrm{E}\left[T_0^{(n)}\right] \geqslant \frac{n}{2}, \quad 其中\ \mu(n) \equiv \max_{0\leqslant m\leqslant n-1} \mathrm{E}[|X_m|].$$

最后, 应用 (a) 得 $\mathrm{E}[T_0] = \infty$. 因此, 根据 (1.2.3) 有 $\mathrm{P}(\rho_0 < \infty) = 1$, 并由此得 $\{X_n : n \geqslant 0\}$ 是常返的.

(c) 为完成证明, 正如在 (1.2.2) 式中的推导那样, 从 (b) 得到

$$\mathrm{P}\left(\rho_0^{(m)} < \infty\right) = 1, \quad m \geqslant 1, \tag{$*$}$$

其中 $\rho_0^{(m)}$ 是第 m 次返回到 0 的时刻. 下面对 $r \in \mathbb{Z}^+$, 令 $\eta_r = \inf\{n \geqslant 0 : X_n \geqslant r\}$, 证明 $\epsilon \equiv \mathrm{P}(\eta_1 > \rho_0) < 1$, 并得出 $\mathrm{P}\left(\eta_1 > \rho_0^{(m)}\right) \leqslant \epsilon^m$. 现在把它和 $(*)$ 式结合起来得 $\mathrm{P}(\eta_1 < \infty) = 1$. 最后, 说明

$$\mathrm{P}(\eta_{r+1} < \infty) \geqslant \mathrm{P}(\eta_r < \infty)\mathrm{P}(\eta_1 < \infty),$$

因此对所有 $r \geqslant 1$, $\mathrm{P}(\eta_r < \infty) = 1$. 因为这意味着对所有 $r \geqslant 1$, $\sup_n X_n \geqslant r$ 以概率 1 成立, 所以 $\sup_n X_n = \infty$ 以概率 1 成立. 为证明 $\inf_n X_n = -\infty$ 以概率 1 成立, 只要用 $\{-X_n : n \geqslant 0\}$ 取代 $\{X_n : n \geqslant 0\}$ 即可.

1.3.12.[11] 以下是一维随机游动在初等排队论的一个有趣的应用. 排队论研究当每个时间段内到达的顾客数和接受服务的顾客数是随机的情况下, 等待服务的顾客数的分布. 这里所考虑的排队模型是最简单的一种. 即假设在时间段 $[n-1,n)$ 内, 到达的顾客数和接受服务的顾客数之差由 $\mathbb{Z}$ 值随机变量 B_n 给定. 进而, 假设 B_n 独立同分布, 满足 $0<\mathrm{E}[|B_n|]<\infty$. 除等待人数不可能为负值外, 所考虑的队列是由 B_n 确定的随机游动 $\{X_n : n \geqslant 0\}$, 其中 $X_0=0, X_n=\sum\limits_{m=1}^{n} B_m$. 排除掉负长度的队列, 排队模型 $\{Q_n : n \geqslant 0\}$ 由下列条件确定

$$Q_0=0, \quad Q_n=(Q_{n-1}+B_n)^+, \quad n \geqslant 1$$

(a) 证明

$$Q_n = X_n - \min_{0\leqslant m\leqslant n} X_m = \max_{0\leqslant m\leqslant n}(X_n - X_m),$$

并证明对每个 $n \geqslant 0$, Q_n 和 $M_n = \max\limits_{0\leqslant m\leqslant n} X_m$ 具有相同的分布.

(b) 令 $M_\infty \equiv \lim\limits_{n\to\infty} M_n \in \mathbb{N}\bigcup\{\infty\}$, 并且由 (a) 得到

$$\lim_{n\to\infty} \mathrm{P}(Q_n = j) = \mathrm{P}(M_\infty = j), \quad j \in \mathbb{N}.$$

(c) 令 $\mu \equiv \mathrm{E}[B_1]$. 由弱大数律知, 对每个 $\epsilon > 0$, 当 $n \to \infty$ 时有 $\mathrm{P}(|X_n - n\mu| \geqslant n\epsilon) \to 0$. 特别地, 当 $\mu > 0$ 时, 证明 $\mathrm{P}(M_\infty = \infty) = 1$. 当 $\mu = 0$ 时, 由习题 1.3.11 可得到相同的结论. 因此, 当 $\mathrm{E}[B_1] \geqslant 0$ 时, 对所有的 $j \in \mathbb{N}$, 有 $\mathrm{P}(Q_n = j) \to 0$. 也就是说, 如果光顾的平均人数不少于所服务的平均人数, 则队列以概率 1 无限增长.

(d) 现在假设 $\mu \equiv \mathrm{E}[B_1] < 0$. 则强大数律 (当 B_1 存在有限四阶矩时, 可参考习题 1.3.4; 而对于一般的情形, 可参考 [9] 中的定理 1.4.11) 告诉我们 $\dfrac{X_n}{n} \to \mu$ 以概率 1 成立. 特别地, 可以推出 $M_\infty < \infty$ 以概率 1 成立, 因此当 $\nu_j \equiv \lim\limits_{n\to\infty} \mathrm{P}(Q_n = j) = \mathrm{P}(M_\infty = j)$ 时, $\sum\limits_{j\in\mathbb{N}} \nu_j = 1$.

(e) 考虑下列特殊情形: 设 B_m 为 $\{-1,1\}$ 值的 Bernoulli 随机变量且 $p \equiv \mathrm{P}(B_1 = 1) \in (0,1)$, 令 $q = 1-p$. 利用 (1.1.12) 式中的计算,

[11] 据我所知, 该例子由 William Feller 设计.

证明

$$\lim_{n\to\infty} P(Q_n = j) = \begin{cases} 0, & 若\ p \geqslant q; \\ \dfrac{q-p}{q}\left(\dfrac{p}{q}\right)^j, & 若\ p < q. \end{cases}$$

(f) 将 (e) 推广到如下情形: $B_m \in \{-1, 0, 1\}$, $p = \mathrm{P}(B_1 = 1)$ 且 $q = \mathrm{P}(B_1 = -1)$. 提示: M_∞ 与 $\sup\limits_n Y_n$ 有相同的分布, 其中 $\{Y_n : n \geqslant 0\}$ 是与 $\{-1,\ 1\}$ 值以概率 $\dfrac{p}{p+q}$ 取 1 的 Bernoulli 随机变量相对应的随机游动.

第二章　Markov 链的 Doeblin 理论

本章正式开始研究 Markov 过程. 像第一章中的随机游动一样, 这里将要研究的随机过程只取可数个值并且有离散的 (与连续相对) 时间参数. 事实上, 这些过程在很多方面都是随机游动最简单的推广. 准确地说, 随机游动增量的分布与增量发生之前的所有事情无关. 我们现在正要研究的随机过程的增量分布只依赖于其当前所处的状态而与其过去所处的状态无关. 具有这种相依性的随机过程被称为具有 *Markov* 性, 并称它为 *Markov* 链[1].

称随机过程取值的集合 $\mathbb{S}$ 为它的*状态空间*, 并且正如前面提到的, 我们所研究的随机过程具有有限或可数的状态空间. 因此, 至少从理论上而言, 没有理由不把 $\mathbb{S}$ 看做集合 $\{1, \cdots, N\}$ 或者 $\mathbb{Z}^+$, 这随 $\mathbb{S}$ 是有限还是可数而定. 另一方面, 总是取 $\mathbb{S}$ 为这两个集合之一也有不利的地方, 因为这样做可能掩盖了一些重要的性质. 比如, 在把 $\mathbb{Z}^2$ 同构映射到 $\mathbb{Z}^+$ 之后, 对 $\mathbb{Z}^2$ 上的最近邻随机游动的描述, 会产生很大的错误.

2.1　概论

在开始之前, 需要了解一些有关 Markov 链的基本知识.

[1] 具有离散时间参数的随机过程常称为"链".

具有有限或可数状态空间 $\mathbb{S}$ 的 *Markov* 链是一族满足下列性质的 $\mathbb{S}$ 值随机变量 $\{X_n : n \geqslant 0\}$: 对所有的 $n \geqslant 0$ 及 $(i_0, \cdots, i_n, j) \in \mathbb{S}^{n+2}$,

$$\mathrm{P}(X_{n+1} = j | X_0 = i_0, \cdots, X_n = i_n) = (\boldsymbol{P})_{i_n j}, \tag{2.1.1}$$

其中 $\boldsymbol{P}$ 是一个所有元素都非负、并且每一行元素之和为 1 的矩阵. 等价地 (参见 §6.4.1)

$$\mathrm{P}(X_{n+1} = j | X_0, \cdots, X_n) = (\boldsymbol{P})_{X_n j}. \tag{2.1.2}$$

应该明确, (2.1.2) 在数学上精确地表达了这样的思想: 当 Markov 链跳跃时, 其到达状态的分布仅仅依赖于跳跃时所处的状态, 与其过去所处的状态无关.

2.1.1. Markov 链的存在性: 很自然地, 具有非负元素且各行元素之和为 1 的矩阵被称为**转移概率矩阵**: 它给出了 Markov 链在 n 时刻处于状态 i 的条件下, $n+1$ 时刻转移到状态 j 的概率, 与其在 n 时刻之前所处的状态无关. 进而, 很明显, 只有转移概率矩阵才能出现在 (2.1.1) 式右端. 可能不那么容易的是, 相反方向的结论也成立. 也就是说, 令 $\boldsymbol{\mu}$ 是一个**概率向量**[2], $\boldsymbol{P}$ 是一个转移概率矩阵. 那么, 存在一个**初始分布为** $\boldsymbol{\mu}$、转移概率矩阵为 $\boldsymbol{P}$ 的 Markov 链 $\{X_n : n \geqslant 0\}$. 也即 $\mathrm{P}(X_0 = i) = (\boldsymbol{\mu})_i$ 且 (2.1.1) 成立.

为了证明上述存在性的结论, 可以采用以下方法. 首先, 不失一般性, 假设 $\mathbb{S}$ 为 $\{1, \cdots, N\}$ 或 $\mathbb{Z}^+$. 然后, 给定 $i \in \mathbb{S}$, 令 $\beta(i, 0) = 0$ 及 $\beta(i, j) = \sum\limits_{k=1}^{j} (\boldsymbol{P})_{ik}$, $j \geqslant 1$, 并且定义 $F : \mathbb{S} \times [0, 1) \to \mathbb{S}$ 使得当 $\beta(i, j-1) \leqslant u < \beta(i, j)$ 时$F(i, u) = j$. 另外, 令 $\alpha(0) = 0$, $\alpha(i) = \sum\limits_{k=1}^{i} (\boldsymbol{\mu})_k$, $i \geqslant 1$, 定义 $f : [0, 1) \to \mathbb{S}$ 使得当 $\alpha(i-1) \leqslant u < \alpha(i)$ 时 $f(u) = i$. 最后, 令 $\{U_n : n \geqslant 0\}$ 为一列相互独立服从 $[0, 1)$ 上均匀分布的随机变量 (参见定理 6.3.2), 令

$$X_n = \begin{cases} f(U_0), & \text{若 } n = 0; \\ F(X_{n-1}, U_n), & \text{若 } n \geqslant 1. \end{cases} \tag{2.1.3}$$

现在证明 (2.1.3) 中的序列 $\{X_n : n \geqslant 0\}$ 是满足以上要求的

[2] 概率向量是行向量, 其各个坐标分量都是非负的并且和为 1.

Markov 链. 为此, 假设 $(i_0, \cdots, i_n) \in \mathbb{S}^{n+1}$, 并注意到

$$\begin{aligned}
&\mathrm{P}(X_0 = i_0, \cdots, X_n = i_n) \\
&= \mathrm{P}(U_0 \in [\alpha(i_0 - 1), \alpha(i_0)), \quad U_m \in [\beta(i_{m-1}, i_m - 1), \beta(i_{m-1}, i_m)), \\
&\quad 1 \leqslant m \leqslant n) \\
&= \boldsymbol{\mu}_{i_0} (\boldsymbol{P})_{i_0 i_1} \cdots (\boldsymbol{P})_{i_{n-1} i_n}.
\end{aligned}$$

2.1.2. 转移概率和概率向量: 这里矩阵记号的使用是明智的. 如果 $\boldsymbol{\mu}$ 是一个行向量, 其第 i 个元素为 $(\boldsymbol{\mu})_i = \mathrm{P}(X_0 = i)$, 则称 $\boldsymbol{\mu}$ 是该马尔可夫链的初始分布, 并且

$$(\boldsymbol{\mu} \boldsymbol{P}^n)_j = \mathrm{P}(X_n = j), \quad n \geqslant 0, \quad j \in \mathbb{S}, \tag{2.1.4}$$

这里我们采用了通常的约定, $\boldsymbol{P}^0$ 为单位矩阵, $\boldsymbol{P}^n = \boldsymbol{P}\boldsymbol{P}^{n-1}$, $n \geqslant 1$.[3] 为了验证 (2.1.4), 给定 $n \geqslant 1$, 注意到根据 (2.1.1) 和归纳法, 有

$$\mathrm{P}(X_0 = i_0, \cdots, X_{n-1} = i_{n-1}, X_n = j) = (\boldsymbol{\mu})_{i_0} (\boldsymbol{P})_{i_0 i_1} \cdots (\boldsymbol{P})_{i_{n-1} j}.$$

因此, 对 $(i_0, \cdots, i_{n-1})$ 求和就可得到 (2.1.4). 显然, (2.1.4)表示当 $\boldsymbol{\mu}$ 是 *Markov* 链的初始分布 (即在 0 时刻的分布) 时, 行向量 $\boldsymbol{\mu}\boldsymbol{P}^n$ 正是其在 n 时刻的分布. 换句话说, $\boldsymbol{P}^n$ 是第 n 步转移概率矩阵: $(\boldsymbol{P}^n)_{ij}$ 是在给定 $X_m = i$ 下, $X_{m+n} = j$ 的条件概率.

为了后面引用的方便, 当行向量被用来表示测度时, 我们这里介绍一种合适的方法来度量它的长度. 即给定一个行向量 $\boldsymbol{\rho}$, 令

$$\|\boldsymbol{\rho}\|_v = \sum_{i \in \mathbb{S}} |(\boldsymbol{\rho})_i|, \tag{2.1.5}$$

其中, 下标 v 表示对应于测度空间上变差范数的长度记号. 选择该范数的基本理由是, 因为根据定理 6.1.15,

$$\|\boldsymbol{\rho}\boldsymbol{P}\|_v = \sum_{j \in \mathbb{S}} \left| \sum_{i \in \mathbb{S}} (\boldsymbol{\rho})_i (\boldsymbol{P})_{ij} \right| \leqslant \sum_{i \in \mathbb{S}} \left(\sum_{j \in \mathbb{S}} |(\boldsymbol{\rho})_i| (\boldsymbol{P})_{ij} \right) = \|\boldsymbol{\rho}\|_v,$$

故有

$$\|\boldsymbol{\rho}\boldsymbol{P}\|_v \leqslant \|\boldsymbol{\rho}\|_v. \tag{2.1.6}$$

[3] 读者可以自行验证对所有 $n \in \mathbb{N}$, $\boldsymbol{P}^n$ 也是一个转移概率矩阵: 各个元素都是非负的且每行元素之和为 1.

要注意的是, 这与 Euclid 度量长度的方法完全不同: Euclid 使用

$$\|\boldsymbol{\rho}\|_2 = \left(\sum_{i\in\mathbb{S}}(\boldsymbol{\rho})_i^2\right)^{\frac{1}{2}}. \tag{2.1.7}$$

另一方面, 当 $\mathbb{S}$ 是有限集时, 这两个范数具有可比性. 即

$$\|\boldsymbol{\rho}\|_2 \leqslant \|\boldsymbol{\rho}\|_v \leqslant \sqrt{\#\mathbb{S}}\|\boldsymbol{\rho}\|_2,$$

其中 $\#\mathbb{S}$ 表示集合 $\mathbb{S}$ 的个数. 第一个不等式通过两边平方很容易得到, 而第二个不等式是 Schwarz 不等式 (参看习题 1.3.1) 的一个应用. 进而, $\|\cdot\|_v$ 在下面意义下是一个好的*范数*(即长度的度量): $\|\boldsymbol{\rho}\|_v = 0$ 当且仅当 $\boldsymbol{\rho} = \mathbf{0}$, 且满足*三角不等式* $\|\boldsymbol{\rho}+\boldsymbol{\rho}'\|_v \leqslant \|\boldsymbol{\rho}\|_v + \|\boldsymbol{\rho}'\|_v$. 最后, Cauchy 收敛准则对 $\|\cdot\|_v$ 成立. 也就是说, 如果 $\{\boldsymbol{\rho}_n\}_1^\infty$ 是 $\mathbb{R}^{\mathbb{S}}$ 中的一个序列, 那么存在 $\boldsymbol{\rho}\in\mathbb{R}^{\mathbb{S}}$ 使得 $\|\boldsymbol{\rho}_n - \boldsymbol{\rho}\|_v \to 0$ 当且仅当 $\{\boldsymbol{\rho}_n\}_1^\infty$ 是 *Cauchy 收敛的*:

$$\lim_{m\to\infty}\sup_{n>m}\|\boldsymbol{\rho}_n - \boldsymbol{\rho}_m\|_v = 0.$$

如通常那样, “仅当” 部分只要应用三角不等式即可:

$$\|\boldsymbol{\rho}_n - \boldsymbol{\rho}_m\|_v \leqslant \|\boldsymbol{\rho}_n - \boldsymbol{\rho}\|_v + \|\boldsymbol{\rho} - \boldsymbol{\rho}_m\|_v.$$

为了证明反方向的结论, 假设 $\{\boldsymbol{\rho}_n\}_1^\infty$ 是 Cauchy 收敛的, 注意到 $\{\boldsymbol{\rho}_n\}_1^\infty$ 的每个分量作为实数一定是 Cauchy 收敛的. 因此, 由关于实数的 Cauchy 收敛准则, 存在 $\boldsymbol{\rho}$ 使得 $\{\boldsymbol{\rho}_n\}_1^\infty$ 收敛于 $\boldsymbol{\rho}$: $\boldsymbol{\rho}_n$ 的每个分量都收敛到 $\boldsymbol{\rho}$ 的相应分量. 从而, 由 Fatou 引理, 定理 6.1.10, 当 $m\to\infty$ 时有

$$\|\boldsymbol{\rho} - \boldsymbol{\rho}_m\|_v = \sum_{i\in\mathbb{S}}|(\boldsymbol{\rho})_i - (\boldsymbol{\rho}_m)_i| \leqslant \liminf_{n\to\infty}\sum_{i\in\mathbb{S}}|(\boldsymbol{\rho}_n)_i - (\boldsymbol{\rho}_m)_i| \to 0.$$

2.1.3. 转移概率和转移函数: 在 §2.1.2 中我们知道, 用矩阵表示转移概率以及用行向量表示初始分布为以后各时刻概率分布的表示提供了方便. 为了理解这种表示对求函数期望值的便利, 我们把状态空间 $\mathbb{S}$ 上的函数 f 看做列向量 $\boldsymbol{f}$, 其第 j 个坐标是函数 f 在 j 点的值. 显然, 若 $\boldsymbol{\mu}$ 表示 $\{1,\cdots,N\}$ 上概率测度 μ 的行向量, 并且 $\boldsymbol{f}$ 表示一个非负或有界函数 f 的列向量, 那么 $\boldsymbol{\mu f} = \sum_{i\in\mathbb{S}} f(i)\mu(\{i\})$ 是 f 关于 μ 的期

望值. 类似地, 列向量 $\boldsymbol{P}^n\boldsymbol{f}$ 所表示的函数在 i 点的值为 $f(X_n)$ 在给定 $X_0=i$ 下的条件期望. 事实上,

$$\begin{aligned}\mathrm{E}[f(X_n)|X_0=i]&=\sum_{j\in\mathbb{S}}f(j)\mathrm{P}(X_n=j|X_0=i)\\&=\sum_{j\in\mathbb{S}}(\boldsymbol{P}^n)_{ij}(\boldsymbol{f})_j=(\boldsymbol{P}^n\boldsymbol{f})_i.\end{aligned}$$

更一般地, 如果 f 是 $\mathbb{S}$ 上的非负或有界函数, $\boldsymbol{f}$ 是由其所决定的列向量, 那么因为

$$\begin{aligned}&\mathrm{E}[f(X_n)|X_0=i_0,\cdots,X_m=i_m]\\=&\sum_{j\in\mathbb{S}}f(j)\mathrm{P}(X_n=j|X_0=i_0,\cdots,X_m=i_m)\\=&\sum_{j\in\mathbb{S}}f(j)(\boldsymbol{P}^{n-m})_{i_mj}=(\boldsymbol{P}^{n-m}\boldsymbol{f})_{i_m},\end{aligned}$$

对 $0\leqslant m\leqslant n$ 有

$$\begin{aligned}\mathrm{E}[f(X_n)|X_0=i_0,\cdots,X_m=i_m]&=(\boldsymbol{P}^{n-m}\boldsymbol{f})_{i_m},\\ \text{或者等价地,}\ \mathrm{E}[f(X_n)|X_0,\cdots,X_m]&=(\boldsymbol{P}^{n-m}\boldsymbol{f})_{X_m}.\end{aligned}\tag{2.1.8}$$

特别地, 如果 $\boldsymbol{\mu}$ 是 $\{X_n:n\geqslant 0\}$ 的初始分布, 那么由于 $\mathrm{E}[f(X_n)]=\sum_i(\boldsymbol{\mu})_i\cdot\mathrm{E}[f(X_n)|X_0=i]$, 所以

$$\mathrm{E}[f(X_n)]=\boldsymbol{\mu}\boldsymbol{P}^n\boldsymbol{f}.\tag{2.1.9}$$

注意到, 正如当我们用行向量表示测度时, $\|\cdot\|_v$ 是度量行向量长度的一种恰当方式一样, 度量表示函数的列向量的长度的恰当方式是取一致范数 $\|\cdot\|_u$:

$$\|\boldsymbol{f}\|_u=\sup_{j\in\mathbb{S}}|(\boldsymbol{f})_j|.\tag{2.1.10}$$

如此选择范数 $\|\cdot\|_u$ 的理由是, 由于

$$|\boldsymbol{\mu}\boldsymbol{f}|\leqslant\sum_{i\in\mathbb{S}}|(\boldsymbol{\mu})_i||(\boldsymbol{f})_i|\leqslant\|\boldsymbol{f}\|_u\sum_{i\in\mathbb{S}}|(\boldsymbol{\mu})_i|,$$

那么 $|\boldsymbol{\mu}\boldsymbol{f}|\leqslant\|\boldsymbol{\mu}\|_v\|\boldsymbol{f}\|_u$. 特别地, 作为 (2.1.6) 的补充, 我们有

$$\|\boldsymbol{P}\boldsymbol{f}\|_u\leqslant\|\boldsymbol{f}\|_u.\tag{2.1.11}$$

2.1.4. Markov 性: 由定义, 如果 μ 是 $\{X_n : n \geqslant 0\}$ 的初始分布, 那么

$$\mathrm{P}(X_0 = i_0, \cdots, X_n = i_n) = (\boldsymbol{\mu})_{i_0}(\boldsymbol{P})_{i_0 i_1} \cdots (\boldsymbol{P})_{i_{n-1} i_n}. \tag{2.1.12}$$

因此, 若 $m, n \geqslant 1$ 且 $F : \mathbb{S}^{n+1} \to \mathbb{R}$ 是有界或非负的, 则

$$\begin{aligned}
&\mathrm{E}[F(X_m, \cdots, X_{m+n}), X_0 = i_0, \cdots, X_m = i_m] \\
= &\sum_{j_1, \cdots, j_n \in \mathbb{S}} F(i_m, j_1, \cdots, j_n)\boldsymbol{\mu}_{i_0}(\boldsymbol{P})_{i_0 i_1} \cdots (\boldsymbol{P})_{i_{m-1} i_m} \cdot \\
&(\boldsymbol{P})_{i_m j_1} \cdots (\boldsymbol{P})_{j_{n-1} j_n} \\
= &\mathrm{E}[F(X_0, \cdots, X_n) | X_0 = i_m]\mathrm{P}(X_0 = i_0, \cdots, X_m = i_m).
\end{aligned}$$

换句话说, 我们已经证明了以下形式的 Markov 性:

$$\begin{aligned}
&\mathrm{E}[F(X_m, \cdots, X_{m+n}) | X_0 = i_0, \cdots, X_m = i_m] \\
= &\mathrm{E}[F(X_0, \cdots, X_n) | X_0 = i_m].
\end{aligned} \tag{2.1.13}$$

2.2 Doeblin 理论

本节将介绍一个初等但基本的 Doeblin 技巧, 它帮助我们研究 Markov 链的长时分布, 特别是有限状态空间上马尔可夫链的长时分布.

2.2.1. Doeblin 基本定理: 由于各种原因, 关于 Markov 链人们想知道的是一个相当长时间之后的分布情况, 并且至少当状态空间有限时, 认为其分布趋于稳定是合理的. 更准确地说, 如果 Markov 链从某状态 i 经一步到达任何状态 j 的概率是正的, 那么由于其状态空间是有限的, 鸽子洞原理表明 Markov 链将反复经历该状态, 并且一段时间之后, 该链的初始分布就被"遗忘"了. 换句话说, 对于这样的 Markov 链, 我们要说明的是, 当 n 充分大时, $\boldsymbol{\mu P}^n$ 几乎与 $\boldsymbol{\mu}$ 无关. 特别地, 这意味着当 m 充分大时, $\boldsymbol{\mu P}^n = (\boldsymbol{\mu P}^{n-m})\boldsymbol{P}^m$ 几乎等于 $\boldsymbol{\mu P}^m$. 因此, 由 Cauchy 收敛准则, $\boldsymbol{\pi} = \lim\limits_{n\to\infty} \boldsymbol{\mu P}^n$ 存在. 另外, 如果该极限存在, 那么我们有 $\boldsymbol{\pi} = \lim\limits_{n\to\infty} \boldsymbol{\mu P}^{n+1} = \lim\limits_{n\to\infty} (\boldsymbol{\mu P}^n)\boldsymbol{P} = \boldsymbol{\pi P}$. 也就是说, $\boldsymbol{\pi}$ 是 $\boldsymbol{P}$ 的一个特征值为 1 的左特征向量. 很自然地, 称满足 $\boldsymbol{\pi} = \boldsymbol{\pi P}$ 的概率向量 $\boldsymbol{\pi}$ 为转移概率矩阵 $\boldsymbol{P}$ 的平稳分布.

尽管前面讨论的是有限状态空间, 但在某些情况下, 上面的讨论甚至适合于无限状态空间. 也就是说, 如果 Markov 链不管从什么状态出发, 总以正概率访问某个给定状态, 那么正如下面定理所证明的那样, 该链将趋于稳定.

2.2.1 Doeblin 定理. 令 $\boldsymbol{P}$ 是转移概率矩阵, 满足性质: 对某状态 $j_0 \in \mathbb{S}$ 和 $\epsilon > 0$, 对所有 $i \in \mathbb{S}$ 都有 $(\boldsymbol{P})_{ij_0} \geqslant \epsilon$. 那么 $\boldsymbol{P}$ 存在唯一的平稳概率向量 $\boldsymbol{\pi}$ 使得 $(\boldsymbol{\pi})_{j_0} \geqslant \epsilon$, 并且对所有的初始分布 $\boldsymbol{\mu}$, 成立

$$\|\boldsymbol{\mu}\boldsymbol{P}^n - \boldsymbol{\pi}\|_v \leqslant 2(1-\epsilon)^n, \quad n \geqslant 0.$$

证明: 关键在于注意到, 如果一个行向量 $\boldsymbol{\rho} \in \mathbb{R}^{\mathbb{S}}$ 满足 $\|\boldsymbol{\rho}\|_v < \infty$, 那么

$$\begin{gathered}\sum_{j\in\mathbb{S}}(\boldsymbol{\rho}\boldsymbol{P})_j = \sum_{i\in\mathbb{S}}(\boldsymbol{\rho})_i \quad \text{并且} \\ \sum_{i\in\mathbb{S}}(\boldsymbol{\rho})_i = 0 \implies \|\boldsymbol{\rho}\boldsymbol{P}^n\|_v \leqslant (1-\epsilon)^n\|\boldsymbol{\rho}\|_v, \quad n \geqslant 1.\end{gathered} \tag{2.2.2}$$

第一个式子很显然, 因为由定理 6.1.15,

$$\sum_{j\in\mathbb{S}}(\boldsymbol{\rho}\boldsymbol{P})_j = \sum_{j\in\mathbb{S}}\left(\sum_{i\in\mathbb{S}}(\boldsymbol{\rho})_i(\boldsymbol{P})_{ij}\right) = \sum_{i\in\mathbb{S}}\left(\sum_{j\in\mathbb{S}}(\boldsymbol{\rho})_i(\boldsymbol{P})_{ij}\right) = \sum_{i\in\mathbb{S}}(\boldsymbol{\rho})_i.$$

至于第二个式子, 注意到由简单的归纳论证, 只需验证当 $n = 1$ 时成立即可. 下面, 假设 $\sum\limits_i(\boldsymbol{\rho})_i = 0$, 并注意到

$$\begin{aligned}|(\boldsymbol{\rho}\boldsymbol{P})_j| &= \left|\sum_{i\in\mathbb{S}}(\boldsymbol{\rho})_i(\boldsymbol{P})_{ij}\right| = \left|\sum_{i\in\mathbb{S}}(\boldsymbol{\rho})_i((\boldsymbol{P})_{ij} - \epsilon\delta_{j,j_0})\right| \\ &\leqslant \sum_{i\in\mathbb{S}}|(\boldsymbol{\rho})_i|((\boldsymbol{P})_{ij} - \epsilon\delta_{j,j_0}),\end{aligned}$$

因此,

$$\begin{aligned}\|\boldsymbol{\rho}\boldsymbol{P}\|_v &\leqslant \sum_{j\in\mathbb{S}}\left(\sum_{i\in\mathbb{S}}|(\boldsymbol{\rho})_i|((\boldsymbol{P})_{ij} - \epsilon\delta_{j,j_0})\right) \\ &= \sum_{i\in\mathbb{S}}|(\boldsymbol{\rho})_i|\left(\sum_{j\in\mathbb{S}}((\boldsymbol{P})_{ij} - \epsilon\delta_{j,j_0})\right) = (1-\epsilon)\|\boldsymbol{\rho}\|_v.\end{aligned}$$

现在令 $\boldsymbol{\mu}$ 为概率向量, 并令 $\boldsymbol{\mu}_n = \boldsymbol{\mu}\boldsymbol{P}^n$. 那么, 由于 $\boldsymbol{\mu}_n = \boldsymbol{\mu}_{n-m}\boldsymbol{P}^m$ 及 $\sum\limits_i((\boldsymbol{\mu}_{n-m})_i - \boldsymbol{\mu}_i) = 1 - 1 = 0$, 故有

$$\|\boldsymbol{\mu}_n - \boldsymbol{\mu}_m\|_v \leqslant (1-\epsilon)^m\|\boldsymbol{\mu}_{n-m} - \boldsymbol{\mu}\|_v \leqslant 2(1-\epsilon)^m, \quad 1 \leqslant m \leqslant n.$$

因此, $\{\boldsymbol{\mu}_n\}_1^\infty$ 是 Cauchy 收敛的; 并且存在 $\boldsymbol{\pi}$ 使得 $\|\boldsymbol{\mu}_n-\boldsymbol{\pi}\|_v\to 0$. 由于每个 $\boldsymbol{\mu}_n$ 都是概率向量, 显然 $\boldsymbol{\pi}$ 也必定是概率向量. 另外, $\boldsymbol{\pi}=\lim\limits_{n\to\infty}\boldsymbol{\mu}\boldsymbol{P}^{n+1}=\lim\limits_{n\to\infty}(\boldsymbol{\mu}\boldsymbol{P}^n)\boldsymbol{P}=\boldsymbol{\pi}\boldsymbol{P}$, 所以 $\boldsymbol{\pi}$ 是平稳的. 特别地,

$$(\boldsymbol{\pi})_{j_0}=\sum_{i\in\mathbb{S}}(\boldsymbol{\pi})_i(\boldsymbol{P})_{ij_0}\geqslant\epsilon\sum_{i\in\mathbb{S}}(\boldsymbol{\pi})_i=\epsilon.$$

最后, 如果 $\boldsymbol{\nu}$ 是任意概率向量, 那么

$$\|\boldsymbol{\nu}\boldsymbol{P}^m-\boldsymbol{\pi}\|_v=\|(\boldsymbol{\nu}-\boldsymbol{\pi})\boldsymbol{P}^m\|_v\leqslant 2(1-\epsilon)^m.$$

这不仅证明了收敛性, 而且还证明了 $\boldsymbol{\pi}$ 是 $\boldsymbol{P}$ 的唯一平稳概率向量. □

从谱理论角度去理解 Doeblin 定理是很有益的. 作为有界函数空间 (具有有限一致范数的列向量) 上的算子, $\boldsymbol{P}$ 有特征值为 1 的右特征函数 $\mathbf{1}$: $\boldsymbol{P}\mathbf{1}=\mathbf{1}$. 因此, 至少当 $\mathbb{S}$ 有限时, 一般原理表明应该存在一个行向量是 $\boldsymbol{P}$ 的特征值为 1 的左特征向量. 进而, 由于 1 及 $\boldsymbol{P}$ 的元素都是实数, 可以选择该左特征向量具有实分量. 这样, 从谱的角度不难理解, 存在非零行向量 $\boldsymbol{\mu}\in\mathbb{R}^{\mathbb{S}}$ 满足 $\boldsymbol{\mu}\boldsymbol{P}=\boldsymbol{\mu}$. 另一方面, 标准谱理论并不能得出 $\boldsymbol{\mu}$ 的各个分量是非负的, 这正是 Doeblin 定理提供有用信息的第一个地方, 该信息甚至当 $\mathbb{S}$ 有限时也不容易从谱理论获得. 为了解释 Doeblin 定理中的估计, 令 $M_1(\mathbb{S};\mathbb{C})$ 表示满足 $\|\boldsymbol{\nu}\|_v=1$ 的行向量 $\boldsymbol{\nu}\in\mathbb{C}^{\mathbb{S}}$ 所组成的空间. 则对所有 $\boldsymbol{\nu}\in M_1(\mathbb{S};\mathbb{C})$有

$$\|\boldsymbol{\nu}\boldsymbol{P}\|_v\leqslant 1,$$

并且

$$\sup\{|\alpha|:\alpha\in\mathbb{C}\ \text{且存在}\ \boldsymbol{\nu}\in M_1(\mathbb{S},\mathbb{C})\ \text{使}\ \boldsymbol{\nu}\boldsymbol{P}=\alpha\boldsymbol{\nu}\}\leqslant 1.$$

此外, 如果 $\boldsymbol{\nu}\boldsymbol{P}=\alpha\boldsymbol{\nu}$ 对某 $\alpha\neq 1$ 成立, 那么 $\boldsymbol{\nu}\mathbf{1}=\boldsymbol{\nu}(\boldsymbol{P}\mathbf{1})=(\boldsymbol{\nu}\boldsymbol{P})\mathbf{1}=\alpha\boldsymbol{\nu}\mathbf{1}$, 从而 $\boldsymbol{\nu}\mathbf{1}=0$. 这样, 由 (2.2.1) 中的估计可知, $\boldsymbol{P}$ 的所有不等于 1 的特征值其绝对值均被 $1-\epsilon$ 所控制. 也就是说 $\boldsymbol{P}$ 的整个谱都位于复单位圆内, 1 是其简单特征值, 并且其他所有特征值都位于半径为 $1-\epsilon$ 的圆内. 最后需要指出的是, 尽管一般谱理论不能解释 Doeblin 定理, 但应该说有一种谱理论, 即由 Frobenius 创立并由 Kakutani 进一步发展起来的谱理论, 确实包含了 Doeblin 的结果. 有兴趣的读者可以参考文献 [2] 中的第 Ⅷ 章.

2.2.2. 两个推广: 定理 2.2.1 的一个实质上明显的推广可以通过注意到下列事实而得到: 对任意的 $M \geqslant 1$ 及 $\epsilon > 0$,[4]

$$\sup_j \inf_i (\boldsymbol{P}^M)_{ij} \geqslant \epsilon \implies \|\boldsymbol{\mu}\boldsymbol{P}^n - \boldsymbol{\pi}\|_v \leqslant 2(1-\epsilon)^{[\frac{n}{M}]} \tag{2.2.3}$$

对所有概率向量 $\boldsymbol{\mu}$ 和唯一的平稳概率向量 $\boldsymbol{\pi}$ 成立. 为了说明这一点, 令 $\boldsymbol{\pi}$ 是由定理 2.2.1 所确定的 $\boldsymbol{P}^M$ 的平稳概率向量, 并且注意到对任何概率向量 $\boldsymbol{\mu}$ 及任意的 $m \in \mathbb{N}, 0 \leqslant r < M$, 有

$$\|\boldsymbol{\mu}\boldsymbol{P}^{mM+r} - \boldsymbol{\pi}\|_v = \|(\boldsymbol{\mu}\boldsymbol{P}^r - \boldsymbol{\pi})\boldsymbol{P}^{mM}\|_v \leqslant 2(1-\epsilon)^m.$$

因此 (2.2.3) 成立, 并且从 (2.2.3) 出发证明 $\boldsymbol{\pi}$ 是 $\boldsymbol{P}$ 的唯一一个平稳测度的推理过程与定理 2.2.1 的证明中给出的相同.

下面的推广有点不明显. 为了理解它的要点, 请记住下面的例子. 考虑 $\{1,2\}$ 上的转移概率矩阵

$$\boldsymbol{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

显然, 这个两状态的链每一步从一个状态到达另一个状态. 从而, 它必定访问它的所有状态. 另一方面, 它并不满足 (2.2.3) 中的假设条件: 当 $i = j$ 且 n 是奇数时, 或者当 $i \neq j$ 且 n 是偶数时, $(\boldsymbol{P}^n)_{ij} = 0$. 因此, 不难看出 (2.2.3) 中的结论对该 $\boldsymbol{P}$ 不成立. 事实上, 容易验证, 尽管 $\left(\frac{1}{2}, \frac{1}{2}\right)$ 是 $\boldsymbol{P}$ 的唯一平稳概率向量, 但对所有的 $n \geqslant 0$, $\left\|(1,0)\boldsymbol{P}^n - \left(\frac{1}{2}, \frac{1}{2}\right)\right\|_v = 1$ 成立. 正如后面 (参看 §3.1.3) 将看到的那样, 这里遇到的问题是由于仅仅当 n 为偶数时 $(\boldsymbol{P}^n)_{11} > 0$ 才成立这一点造成的.

不管由前面例子提出的问题如何, 人们会期望与 $\boldsymbol{P}$ 所对应的链在某种意义下趋于均衡. 为了说明这一点, 令

$$\boldsymbol{A}_n = \frac{1}{n}\sum_{m=0}^{n-1} \boldsymbol{P}^m. \tag{2.2.4}$$

[4] 这里和其他地方, $[s]$ 表示 $s(s \in \mathbb{R})$ 的整数部分. 即 $[s]$ 是小于或等于 s 的最大整数.

尽管 $\boldsymbol{A}_n$ 是一个转移概率矩阵, 但它并不表示转移, 而是给出了该链访问各状态的平均时间. 准确地说, 由于

$$(\boldsymbol{A}_n)_{ij} = \frac{1}{n}\sum_{m=0}^{n-1} \mathrm{P}(X_m = j|X_0 = i) = \mathrm{E}\left[\frac{1}{n}\sum_{m=0}^{n-1} \mathbf{1}_{\{j\}}(X_m)|X_0 = i\right],$$

因此 $(\boldsymbol{A}_n)_{ij}$ 是该链从初始状态 i 出发, 在时间区间 $[0, n-1]$ 内处于状态 j 的平均时间的期望值. 经验表明当取平均时数据会变得平缓得多, 现在的情况也不例外. 事实上, 继续考虑上面的例子, 注意到对任意概率向量 $\boldsymbol{\mu}$, 有

$$\left\|\boldsymbol{\mu}\boldsymbol{A}_n - \left(\frac{1}{2}, \frac{1}{2}\right)\right\|_v \leqslant \frac{1}{n}, \quad n \geqslant 1.$$

下面的定理表明以上结论具有一般性.

2.2.5 定理. *假设 $\boldsymbol{P}$ 是 $\mathbb{S}$ 上的转移概率矩阵. 如果对某一 $M \in \mathbb{Z}^+$, $j_0 \in \mathbb{S}$ 及 $\epsilon > 0$, $(\boldsymbol{A}_M)_{ij_0} \geqslant \epsilon$ 对所有 $i \in \mathbb{S}$ 成立, 那么, 存在 $\boldsymbol{P}$ 的唯一平稳概率向量 $\boldsymbol{\pi}$, 使得 $(\boldsymbol{\pi})_{j_0} \geqslant \epsilon$, 并且对任意概率向量 $\boldsymbol{\mu}$ 有*

$$\|\boldsymbol{\mu}\boldsymbol{A}_n - \boldsymbol{\pi}\|_v \leqslant \frac{M-1}{n\epsilon}.$$

首先, 令 $\boldsymbol{\pi}$ 是定理 2.2.1 所确定的 $\boldsymbol{A}_M$ 的唯一平稳概率向量. 那么, 由于任意关于 $\boldsymbol{P}$ 平稳的概率向量 $\boldsymbol{\mu}$ 也必定关于 $\boldsymbol{A}_M$ 平稳, 所以很显然, $\boldsymbol{\pi}$ 是唯一可能的 $\boldsymbol{P}$-平稳概率向量. 进而, 为了证明 $\boldsymbol{\pi}$ 是 $\boldsymbol{P}$-平稳的, 注意到由于 $\boldsymbol{P}$ 与 $\boldsymbol{A}_M$ 可交换, $(\boldsymbol{\pi}\boldsymbol{P})\boldsymbol{A}_M = (\boldsymbol{\pi}\boldsymbol{A}_M)\boldsymbol{P} = \boldsymbol{\pi}\boldsymbol{P}$. 因此, $\boldsymbol{\pi}\boldsymbol{P}$ 是关于 $\boldsymbol{A}_M$ 平稳的, 并由唯一性知, $\boldsymbol{\pi}\boldsymbol{P}$ 一定等于 $\boldsymbol{\pi}$. 即 $\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}$.

为了证明上述收敛性结果, 需要一个平均化的基本性质. 即对任意概率向量 $\boldsymbol{\mu}$ 及所有 $m, n \geqslant 1$, 有

$$\|\boldsymbol{\mu}\boldsymbol{A}_n\boldsymbol{A}_m - \boldsymbol{\mu}\boldsymbol{A}_n\|_v \leqslant \frac{m-1}{n}. \tag{2.2.6}$$

为了验证这一性质, 首先注意到, 由三角不等式得

$$\begin{aligned}\|\boldsymbol{\mu}\boldsymbol{A}_n\boldsymbol{A}_m - \boldsymbol{\mu}\boldsymbol{A}_n\|_v &= \frac{1}{m}\left\|\sum_{k=0}^{m-1}(\boldsymbol{\mu}\boldsymbol{A}_n\boldsymbol{P}^k - \boldsymbol{\mu}\boldsymbol{A}_n)\right\|_v \\ &\leqslant \frac{1}{m}\sum_{k=0}^{m-1}\|\boldsymbol{\mu}\boldsymbol{A}_n\boldsymbol{P}^k - \boldsymbol{\mu}\boldsymbol{A}_n\|_v.\end{aligned}$$

其次, 对每个 $k \geqslant 0$,

$$\boldsymbol{\mu}\boldsymbol{A}_n\boldsymbol{P}^k - \boldsymbol{\mu}\boldsymbol{A}_n = \frac{1}{n}\sum_{\ell=0}^{n-1}(\boldsymbol{\mu}\boldsymbol{P}^{\ell+k} - \boldsymbol{\mu}\boldsymbol{P}^\ell) = \frac{1}{n}\left(\sum_{\ell=k}^{n+k-1}\boldsymbol{\mu}\boldsymbol{P}^\ell - \sum_{\ell=0}^{n-1}\boldsymbol{\mu}\boldsymbol{P}^\ell\right),$$

因此 $\|\boldsymbol{\mu A}_n\boldsymbol{P}^k - \boldsymbol{\mu A}_n\|_v \leqslant \frac{2k}{n}$. 再结合上面的不等式, 得到

$$\|\boldsymbol{\mu A}_n\boldsymbol{A}_m - \boldsymbol{\mu A}_n\|_v \leqslant \frac{2}{mn}\sum_{k=0}^{m-1} k = \frac{m-1}{n},$$

这正是我们想要的结论.

为了从这儿出发完成定理 2.2.5 的证明, 假设对所有的 i, $(\boldsymbol{A}_M)_{ij_0} \geqslant \epsilon$, 并且和上面一样, 令 $\boldsymbol{\pi}$ 是 $\boldsymbol{P}$ 的唯一平稳概率向量. 则 $\boldsymbol{\pi}$ 也是 $\boldsymbol{A}_M$ 的唯一平稳概率向量, 因此将 (2.2.2) 中第二行的估计应用到 $\boldsymbol{A}_M$, 得到 $\|\boldsymbol{\mu A}_n\boldsymbol{A}_M - \boldsymbol{\pi}\|_v = \|(\boldsymbol{\mu A}_n - \boldsymbol{\pi})\boldsymbol{A}_M\|_v \leqslant (1-\epsilon)\|\boldsymbol{\mu A}_n - \boldsymbol{\pi}\|_v$, 它和 (2.2.6) 相结合, 给出

$$\begin{aligned}\|\boldsymbol{\mu A}_n - \boldsymbol{\pi}\|_v &\leqslant \|\boldsymbol{\mu A}_n - \boldsymbol{\mu A}_n\boldsymbol{A}_M\|_v + \|\boldsymbol{\mu A}_n\boldsymbol{A}_M - \boldsymbol{\pi}\|_v \\ &\leqslant \frac{M-1}{n} + (1-\epsilon)\|\boldsymbol{\mu A}_n - \boldsymbol{\pi}\|_v.\end{aligned}$$

最后, 经过初等整理, 给出所要的结果.

2.3 遍历理论要素

在前一节, 我们知道在适当的条件下, $\boldsymbol{\mu P}^n$ 或 $\boldsymbol{\mu A}_n$ 收敛, 并且极限是 $\boldsymbol{P}$ 的唯一平稳概率向量 $\boldsymbol{\pi}$. 在这一节, 我们将给出这些结果的更具有概率论特色的解释. 特别地, 我们将给出 $\boldsymbol{\pi}$ 的概率解释. 第三章将采用完全不同的方法加以解释.

在进一步分析之前, 有必要将以前的结果概括如下 (参见 (2.2.3) 并记住 $|\boldsymbol{\mu f}| \leqslant \|\boldsymbol{\mu}\|_v\|\boldsymbol{f}\|_u$): [5]

$$\sup_j \inf_i (\boldsymbol{P}^M)_{ij} \geqslant \epsilon \implies \|\boldsymbol{Pf} - \boldsymbol{\pi f}\|_u \leqslant 2(1-\epsilon)^{\left[\frac{n}{M}\right]}\|\boldsymbol{f}\|_u \qquad (2.3.1)$$

和 (参看定理 2.2.5)

$$\sup_j \inf_i (\boldsymbol{A}_M)_{ij} \geqslant \epsilon \implies \|\boldsymbol{A}_n\boldsymbol{f} - \boldsymbol{\pi f}\|_u \leqslant \frac{M-1}{n\epsilon}\|\boldsymbol{f}\|_u, \qquad (2.3.2)$$

其中 $\boldsymbol{f}$ 是有界列向量.

2.3.1. 平均遍历定理: 令 $\{X_n : n \geqslant 0\}$ 是具有转移概率 $\boldsymbol{P}$ 的 Markov 链. 显然,

$$\overline{T}_j^{(n)} \equiv \frac{1}{n}\sum_{m=0}^{n-1} \mathbf{1}_{\{j\}}(X_m) \qquad (2.3.3)$$

[5] 在本书中, 我们将用常数来表示对应的常数函数.

是该 Markov 链在 n 时刻之前处于状态 j 的平均时间. 这样, 如果 $\boldsymbol{\mu}$ 是其初始分布 (即 $(\boldsymbol{\mu})_i = \mathrm{P}(X_0 = i)$), 那么 $(\boldsymbol{\mu}\boldsymbol{A}_n)_j = \mathrm{E}\left[\overline{T}_j^{(n)}\right]$, 由此并利用定理 2.2.5 知, 当 $n \to \infty$ 时, $\mathrm{E}[\overline{T}_j^{(n)}] \to (\boldsymbol{\pi})_j$. 我们将证明当 $n \to \infty$ 时, 不仅数学期望, 而且随机变量 $\overline{T}_j^{(n)}$ 本身收敛到 $(\boldsymbol{\pi})_j$. 这类结果属于*遍历理论*范畴. 遍历理论是一个数学原理, 最早由物理学家 Gibbs 在研究气体分子动力学理论时提出, 它断言一个随机动力系统关于某给定轨道的时间平均和该系统的均衡状态近似相同. 不幸的是, 尽管像这里给出的这些结果证实了该原理, 但甚至在 Gibbs 之后差不多 150 年的今天, 本质上也不存在使 Gibbs 原理得到数学证明而物理上可实现的情形.

2.3.4 平均遍历定理. *在定理 2.2.5 的假设条件下, 对所有 $n \geqslant 1$, 有*

$$\sup_{j\in\mathbb{S}} \mathrm{E}\left[\left(\overline{T}_j^{(n)} - (\boldsymbol{\pi})_j\right)^2\right] \leqslant \frac{2(M-1)}{n\epsilon}.$$

(对于更精确的非定量化的形式参看下文的 (2.3.10).) 更一般地, 对 $\mathbb{S}$ 上任意有界函数 f 及所有 $n \geqslant 1$, 有

$$\mathrm{E}\left[\left(\frac{1}{n}\sum_{m=0}^{n-1} f(X_m) - \boldsymbol{\pi}\boldsymbol{f}\right)^2\right] \leqslant \frac{2(M-1)\|\boldsymbol{f}\|_u^2}{n\epsilon},$$

其中 $\boldsymbol{f}$ 表示由 f 所确定的列向量.

证明: 令 $\overline{\boldsymbol{f}}$ 是由函数 $\overline{f} = f - \boldsymbol{\pi}\boldsymbol{f}$ 所确定的列向量. 显然,

$$\frac{1}{n}\sum_{m=0}^{n-1} f(X_m) - \boldsymbol{\pi}\boldsymbol{f} = \frac{1}{n}\sum_{m=0}^{n-1} \overline{f}(X_m),$$

并且因此

$$\begin{aligned}\left(\frac{1}{n}\sum_{m=0}^{n-1} f(X_m) - \boldsymbol{\pi}\boldsymbol{f}\right)^2 &= \frac{1}{n^2}\left(\sum_{m=0}^{n-1}\overline{f}(X_m)\right)^2 = \frac{1}{n^2}\sum_{k,\ell=0}^{n-1}\overline{f}(X_k)\overline{f}(X_\ell)\\ &= \frac{2}{n^2}\sum_{0\leqslant k\leqslant \ell<n}\overline{f}(X_k)\overline{f}(X_\ell) - \frac{1}{n^2}\sum_{k=0}^{n-1}\overline{f}(X_k)^2\\ &\leqslant \frac{2}{n^2}\sum_{0\leqslant k\leqslant \ell<n}\overline{f}(X_k)\overline{f}(X_\ell).\end{aligned}$$

所以

$$\begin{aligned}\mathrm{E}\left[\left(\frac{1}{n}\sum_{m=0}^{n-1}f(X_m)-\boldsymbol{\pi f}\right)^2\right] &\leqslant \frac{2}{n^2}\sum_{k=0}^{n-1}\mathrm{E}\left[\overline{f}(X_k)\sum_{\ell=0}^{n-k-1}\overline{f}(X_{k+\ell})\right]\\ &=\frac{2}{n^2}\sum_{k=0}^{n-1}\mathrm{E}\left[\overline{f}(X_k)\sum_{\ell=0}^{n-k-1}(\boldsymbol{P}^{\ell}\overline{\boldsymbol{f}})_{X_k}\right]\\ &=\frac{2}{n^2}\sum_{k=0}^{n-1}(n-k)\mathrm{E}\left[\overline{f}(X_k)(\boldsymbol{A}_{n-k}\overline{\boldsymbol{f}})_{X_k}\right].\end{aligned}$$

但是, 由 (2.3.2) 知, $\|\boldsymbol{A}_{n-k}\overline{\boldsymbol{f}}\|_u \leqslant \dfrac{M-1}{(n-k)\epsilon}\|\overline{\boldsymbol{f}}\|_u$, 并且因为 $\|\overline{\boldsymbol{f}}\|_u \leqslant \|\boldsymbol{f}\|_u$, 我们有

$$(n-k)\mathrm{E}\left[\overline{f}(X_k)(\boldsymbol{A}_{n-k}\overline{\boldsymbol{f}})_{X_k}\right] \leqslant \frac{(M-1)\|\boldsymbol{f}\|_u^2}{\epsilon}.$$

将此式代入上式, 就得到了定理的第二个结论. 为了证明第一个结论, 只要取 $f=\mathbf{1}_{\{j\}}$ 并注意到此时 $\|\boldsymbol{f}\|_u\leqslant 1$ 即可. □

2.3.2. 返回次数: 正如 §1.1 和 §1.2 中所指出的那样, 返回次数在 Markov 链长时性质的分析中应该起着重要的作用. 特别地, 如果 $\rho_j^{(0)}\equiv 0$, 并且对 $m\geqslant 1$, **第 m 次返回到状态 j 的时刻** $\rho_j^{(m)}$ 定义如下: 若 $\rho_j^{(m-1)}=\infty$, 或对每一 $n>\rho_{j-1}^{(m-1)}$, $X_n\neq j$, 则定义 $\rho_j^{(m)}=\infty$, 否则$\rho_j^{(m)}=\inf\{n>\rho_j^{(m-1)}:\ X_n=j\}$. 那么, 根据 $\mathrm{P}(\rho_j^{(1)}<\infty|X_0=j)$ 等于 1 或小于 1, 我们称 j 是**常返的**或**瞬时的**; 并且希望当 j 是常返状态时, 该 Markov 链的时间域被状态 j 的相继返回时刻分成各个不同的时期. 在这一小节, 我们将提供一些证据支持这一点.

注意到 $\rho_j\equiv\rho_j^{(1)}\geqslant 1$, 并且对 $n\geqslant 1$,

$$\begin{aligned}&\mathbf{1}_{(n,\infty]}(\rho_j)=F_{n,j}(X_0,\cdots,X_n),\\ &\text{其中 } F_{n,j}(i_0,\cdots,i_n)=\begin{cases}1, & \text{若对 } 1\leqslant m\leqslant n,\ i_m\neq j;\\ 0, & \text{其他.}\end{cases}\end{aligned}\tag{2.3.5}$$

特别地, 这表明事件 $\{\rho_j>n\}$ 是 $(X_0,\cdots,X_n)$ 的可测函数. 更一般地, 由于

$$\mathbf{1}_{(n,\infty]}\left(\rho_j^{(m+1)}\right)=\mathbf{1}_{[n,\infty]}\left(\rho_j^{(m)}\right)+\sum_{\ell=1}^{n-1}\mathbf{1}_{\{\ell\}}\left(\rho_j^{(m)}\right)F_{n-\ell,j}(X_\ell,\cdots,X_n),$$

简单的归纳证明可以得出, 对每个 $m \in \mathbb{N}$ 及 $n \in \mathbb{N}$, $\{\rho_j^{(m)} > n\}$ 是 $(X_0, \cdots, X_n)$ 的可测函数.

2.3.6 定理. 对所有的 $m \in \mathbb{Z}^+$ 及 $(i,j) \in \mathbb{S}^2$,

$$\mathrm{P}\left(\rho_j^{(m)} < \infty | X_0 = i\right) = \mathrm{P}(\rho_j < \infty | X_0 = i)\mathrm{P}(\rho_j < \infty | X_0 = j)^{m-1}.$$

特别地, 如果 j 是常返的, 那么对所有的 $m \in \mathbb{N}$, $\mathrm{P}(\rho_j^{(m)} < \infty | X_0 = j) = 1$. 事实上, 如果 j 是常返的, 那么在 $X_0 = j$ 的条件下, $\{\rho_j^{(m)} - \rho_j^{(m-1)} : m \geqslant 1\}$ 是一列相互独立的随机变量, 每一个都和 ρ_j 有相同的分布.

证明: 为证明第一个结论, 应用 (2.1.13) 式及单调收敛定理, 定理 6.1.9 得到

$$\begin{aligned}
&\mathrm{P}\left(\rho_j^{(m)} < \infty | X_0 = i\right) \\
&= \sum_{n=1}^{\infty} \mathrm{P}\left(\rho_j^{(m-1)} = n, \rho_j^{(m)} < \infty | X_0 = i\right) \\
&= \sum_{n=1}^{\infty} \lim_{N \to \infty} \mathrm{E}\left[1 - F_{N,j}(X_n, \cdots, X_{n+N}),\ \rho_j^{(m-1)} = n | X_0 = i\right] \\
&= \sum_{n=1}^{\infty} \lim_{N \to \infty} \mathrm{E}[1 - F_{N,j}(X_0, \cdots, X_N) | X_0 = i]\mathrm{P}\left(\rho_j^{(m-1)} = n | X_0 = i\right) \\
&= \sum_{n=1}^{\infty} \lim_{N \to \infty} \mathrm{P}(\rho_j \leqslant N | X_0 = j)\mathrm{P}\left(\rho_j^{(m-1)} = n | X_0 = i\right) \\
&= \mathrm{P}(\rho_j < \infty | X_0 = j)\mathrm{P}\left(\rho_j^{(m-1)} < \infty | X_0 = i\right).
\end{aligned}$$

至于第二个结论, 只需证明

$$\begin{aligned}
&\mathrm{P}\left(\rho_j^{(m+1)} > n + n_m | X_0 = j,\ \rho_j^{(1)} = n_1, \cdots, \rho_j^{(m)} = n_m\right) \\
&= \mathrm{P}(\rho_j > n | X_0 = j).
\end{aligned}$$

但是, 再次利用 (2.1.13) 式, 左边等于

$$\begin{aligned}
&\mathrm{E}\left[F_{n,j}(X_{n_m}, \cdots, X_{n_m+n}) | X_0 = j, \rho_j^{(1)} = n_1, \cdots, \rho_j^{(m)} = n_m\right] \\
&= \mathrm{E}[F_{n,j}(X_0, \cdots, X_n) | X_0 = j] = \mathrm{P}(\rho_j > n | X_0 = j). \quad \square
\end{aligned}$$

正如 §1.2.2 中的推导一样, 从定理 2.3.6 的第一部分可得:

$$\begin{aligned}
&\mathrm{E}[T_j|X_0=i]=\delta_{i,j}+\frac{\mathrm{P}(\rho_j<\infty|X_0=i)}{\mathrm{P}(\rho_j=\infty|X_0=j)},\\
&\mathrm{E}[T_j|X_0=j]=\infty \Longleftrightarrow \mathrm{P}(T_j=\infty|X_0=j)=1,\\
&\mathrm{E}[T_j|X_0=j]<\infty \Longleftrightarrow \mathrm{P}(T_j<\infty|X_0=j)=1,
\end{aligned}\tag{2.3.7}$$

其中 $T_j=\sum\limits_{m=0}^{\infty}\mathbf{1}_{\{j\}}(X_m)$ 是该链处于状态 j 的总时间. 事实上, 因为

$$\mathrm{P}(T_j>m|X_0=i)=\begin{cases}\mathrm{P}\left(\rho_j^{(m)}<\infty|X_0=j\right), & 若\ i=j;\\ \mathrm{P}\left(\rho_j^{(m+1)}<\infty|X_0=i\right), & 若\ i\neq j.\end{cases}$$

所以, 从定理 2.3.6 的第一部分马上可以得到 (2.3.7) 的三个关系式.

当然, 由 (2.3.7) 式可知, j 是常返的当且仅当 $\mathrm{E}[T_j|X_0=j]=\infty$. 特别地, 在定理 2.2.5 的条件下, 有 $(\boldsymbol{A}_n)_{j_0j_0}\to(\boldsymbol{\pi})_{j_0}>0$, 并且因此

$$\mathrm{E}[T_{j_0}|X_0=j_0]=\sum_{m=0}^{\infty}(\boldsymbol{P}^m)_{j_0j_0}=\lim_{n\to\infty}n(\boldsymbol{A}_n)_{j_0j_0}=\infty.$$

也就是说, 定理 2.2.5 中的条件推出 j_0 是常返的, 并且正如下面将要证明的那样, 从这些条件可以得到更多的结果.

为了便于叙述下一个结果, 如果存在 $n\geqslant 0$ 使得 $(\boldsymbol{P}^n)_{ij}>0$, 那么称 j 从 i 是可到达的, 并记作 $i\to j$. 等价地, $i\to j$ 当且仅当 $i=j$ 或 $\mathrm{P}(\rho_j<\infty|X_0=i)>0$.

2.3.8 定理. 假设对某 $M\geqslant 1$, j_0 及 $\epsilon>0$, 有 $\inf\limits_i(\boldsymbol{A}_M)_{ij_0}\geqslant\epsilon$. 则 j 是常返的当且仅当 $j_0\to j$. 进而, 若 $j_0\to j$, 则对所有 $p\in(0,\infty)$, 有 $\mathrm{E}[\rho_j^p|X_0=j]<\infty$.

证明: 首先假设 $j_0\nrightarrow j$. 等价地, $\mathrm{P}(\rho_j=\infty|X_0=j_0)=1$. 同时, 由于 $(\boldsymbol{A}_M)_{jj_0}\geqslant\epsilon$, 那么存在 $1\leqslant m<M$, 使得 $(\boldsymbol{P}^m)_{jj_0}>0$. 因此

$$\begin{aligned}
&\mathrm{P}(\rho_j^{(m)}=\infty|X_0=j)\\
\geqslant\ &\mathrm{P}(\rho_j^{(m)}=\infty, X_m=j_0|X_0=j)\\
\geqslant\ &\lim_{N\to\infty}\mathrm{E}[F_{N,j}(X_m,\cdots,X_{m+N}),\ X_m=j_0|X_0=j]\\
=\ &\lim_{N\to\infty}\mathrm{E}[F_{N,j}(X_0,\cdots,X_N)|X_0=j_0]\mathrm{P}(X_m=j_0|X_0=j)\\
=\ &\mathrm{E}(\rho_j=\infty|X_0=j_0)(\boldsymbol{P}^m)_{jj_0}>0.
\end{aligned}$$

所以, 由定理 2.3.6 知 j 不可能是常返的.

接下来证明

$$j_0 \to j \implies \text{对某个 } M' \geqslant 1, \inf_i (\boldsymbol{A}_{M'})_{ij} > 0. \tag{$*$}$$

为此, 选择 $m \in \mathbb{N}$ 使得 $(\boldsymbol{P}^m)_{j_0 j} > 0$. 则对所有 $i \in \mathbb{S}$, 有

$$\begin{aligned}(\boldsymbol{A}_{m+M})_{ij} &= \frac{1}{m+M}\sum_{l=0}^{M+m-1}(\boldsymbol{P}^l)_{ij} \geqslant \frac{1}{m+M}\sum_{l=0}^{M-1}(\boldsymbol{P}^l)_{ij_0}(\boldsymbol{P}^m)_{j_0 j}\\ &= \frac{M}{m+M}(\boldsymbol{A}_M)_{ij_0}(\boldsymbol{P}^m)_{j_0 j} \geqslant \frac{M\epsilon}{m+M}(\boldsymbol{P}^m)_{j_0 j} > 0.\end{aligned}$$

根据 $(*)$ 和已经证明过的结论, 只要证明: 如果对某 $\epsilon > 0$ 和 $M \in \mathbb{Z}^+$ 有 $\inf_i (\boldsymbol{A}_M)_{ij} \geqslant \epsilon$, 那么 $\mathrm{E}[\rho_j^p | X_0 = j] < \infty$. 为此, 对 $n \in \mathbb{Z}^+$, $i \in \mathbb{S}$, 令 $u(n,i) = \mathrm{P}(\rho_j > nM | X_0 = i)$. 那么由 (2.1.13) 式, 有

$$\begin{aligned}&u(n+1, i)\\ &= \sum_{k \neq j}\mathrm{P}(\rho_j > (n+1)M,\ X_{nM} = k | X_0 = i)\\ &= \sum_{k \neq j}\mathrm{E}[F_{M,j}(X_{nM}, \cdots, X_{(n+1)M}),\ \rho_j > nM,\ X_{nM} = k | X_0 = i]\\ &= \sum_{k \neq j}\mathrm{P}(\rho_j > M | X_0 = k)\mathrm{P}(\rho_j > nM,\ X_{nM} = k | X_0 = i)\\ &= \sum_{k \neq j}u(1,k)\mathrm{P}(\rho_j > nM,\ X_{nM} = k | X_0 = i).\end{aligned}$$

因此, $u(n+1,i) \leqslant Uu(n,i)$, 其中 $U \equiv \max_{k \neq j} u(1,k)$. 最后, 由于 $u(1,k) = 1 - \mathrm{P}(\rho_j \leqslant M | X_0 = k)$ 和

$$\mathrm{P}(\rho_j \leqslant M | X_0 = k) \geqslant \max_{1 \leqslant m < M}(\boldsymbol{P}^m)_{kj} \geqslant (\boldsymbol{A}_M)_{kj} \geqslant \epsilon, \quad k \neq j,$$

得 $U \leqslant 1 - \epsilon$. 特别地, 这意味着 $u(n+1,j) \leqslant (1-\epsilon)u(n,j)$, 因此 $\mathrm{P}(\rho_j > nM | X_0 = j) \leqslant (1-\epsilon)^n$. 由此立即得到

$$\begin{aligned}\mathrm{E}[\rho_j^p | X_0 = j] &= \sum_{n=1}^{\infty} n^p \mathrm{P}(\rho_j = n | X_0 = j)\\ &\leqslant \sum_{m=1}^{\infty}(mM)^p \sum_{n=(m-1)M+1}^{mM}\mathrm{P}(\rho_j = n | X_0 = j)\end{aligned}$$

$$\leqslant M^p \sum_{m=1}^{\infty} m^p \mathrm{P}(\rho_j > (m-1)M | X_0 = j)$$
$$\leqslant M^p \sum_{m=1}^{\infty} m^p (1-\epsilon)^{m-1} < \infty. \ \square$$

2.3.3. π 的确定: 在定理 2.2.5 的条件下, 存在一个 $\boldsymbol{P}$-平稳概率向量 $\boldsymbol{\pi}$. 在这一节里, 我们将给出 $(\boldsymbol{\pi})_j$ 的一个概率解释, 也即证明

$$\begin{aligned} &\sup_{M\geqslant 1}\sup_{j\in\mathbb{S}}\inf_{i\in\mathbb{S}}(\boldsymbol{A}_M)_{ij} > 0 \\ \Longrightarrow\ &(\boldsymbol{\pi})_j = \frac{1}{\mathrm{E}[\rho_j|X_0=j]} \quad (\equiv 0, \text{如果 } j \text{ 是暂留的}). \end{aligned} \tag{2.3.9}$$

(2.3.9) 式的证明思路如下. 一方面 (参看 (2.3.3)),

$$\mathrm{E}[\overline{T}_j^{(n)}|X_0=j] = (\boldsymbol{A}_n)_{jj} \to (\boldsymbol{\pi})_j,$$

另一方面,

$$X_0 = j \implies \overline{T}_j^{(\rho_j^{(m)})} = \frac{1}{\rho_j^{(m)}} \sum_{\ell=0}^{\rho_j^{(m)}-1} \mathbf{1}_{\{j\}}(X_\ell) = \frac{m}{\rho_j^{(m)}}.$$

这样, 因为 $\rho_j^{(m)}$ 是 ρ_j 的 m 个独立复制之和, 所以和弱大数律结合起来, 可以得到

$$(\boldsymbol{\pi})_j = \lim_{m\to\infty} \mathrm{E}\left[\overline{T}_j^{(\rho_j^{(m)})} | X_0 = j\right] = \frac{1}{\mathrm{E}[\rho_j|X_0=j]}.$$

按照上面的思想, 实际上可以证明一个更强的结论. 即对每个 $j \in \mathbb{S}$,[6]

$$\mathrm{P}\left(\lim_{n\to\infty} \overline{T}_j^{(n)} = \frac{1}{\mathrm{E}[\rho_j|X_0=j]} \,\middle|\, X_0 = j\right) = 1. \tag{2.3.10}$$

特别地, 因为 $0 \leqslant \overline{T}_j^{(n)} \leqslant 1$, Lebesgue 控制收敛定理和定理 6.1.11 表明, 从 (2.3.10) 式可以推出

$$(\boldsymbol{\pi})_j = \lim_{n\to\infty} (\boldsymbol{A}_n)_{jj} = \lim_{n\to\infty} \mathrm{E}[\overline{T}_j^{(n)}|X_0=j] = \frac{1}{\mathrm{E}[\rho_j|X_0=j]}.$$

[6] 类似于下面的结论被称作逐点遍历定理, 因为与定理 2.3.4 的第一部分不同, 它们是关于以概率 1 逐点收敛的, 而不是关于平均收敛的. 更多内容可参看习题 3.3.9.

这样, 只需证明 (2.3.10) 式即可. 为此, 选择 j_0, M 和 $\epsilon > 0$, 使得对于所有的 i, $(\boldsymbol{A}_M)_{ij_0} \geqslant \epsilon$. 如果 $j_0 \nrightarrow j$, 那么由定理 2.3.8, j 是瞬时的, 并且由 (2.3.7) 式, $\mathrm{P}(T_j < \infty | X_0 = j) = 1$. 因此, 在 $X_0 = j$ 的条件下, $\overline{T}_j^{(n)} \leqslant \frac{1}{n} T_j \to 0$ 以概率 1 成立. 同时, 因为 j 是瞬时的, 所以 $\mathrm{P}(\rho_j = \infty | X_0 = j) > 0$, 因而 $\mathrm{E}[\rho_j | X_0 = j] = \infty$. 因此, 证明了当 $j_0 \nrightarrow j$ 时, (2.3.10) 式成立.

下面假设 $j_0 \to j$. 那么, 再次利用定理 2.3.8, $\mathrm{E}[\rho_j^4 | X_0 = j] < \infty$, 并且在 $X_0 = j$ 的条件下, $\{\rho_j^{(m)} - \rho_j^{(m-1)} : m \geqslant 1\}$ 是一列相互独立的随机变量序列, 并和 ρ_j 有相同的分布. 特别地, 由强大数律 (参看习题 1.3.4)

$$\mathrm{P}\left(\lim_{m \to \infty} \frac{\rho_j^{(m)}}{m} = r_j \,\middle|\, X_0 = j \right) = 1, \quad \text{其中 } r_j \equiv \mathrm{E}[\rho_j | X_0 = j].$$

另一方面, 对任意 $m \geqslant 1$,

$$|\overline{T}_j^{(n)} - r_j^{-1}| \leqslant |\overline{T}_j^{(n)} - \overline{T}_j^{(\rho_j^{(m)})}| + |\overline{T}_j^{(\rho_j^{(m)})} - r_j^{-1}|,$$

且

$$\begin{aligned}\left| \overline{T}_j^{(n)} - \overline{T}_j^{(\rho_j^{(m)})} \right| &\leqslant \frac{|T_j^{(n)} - T_j^{(\rho_j^{(m)})}|}{n} + \left| 1 - \frac{\rho_j^{(m)}}{n} \right| \overline{T}_j^{(\rho_j^{(m)})} \\ &\leqslant 2 \left| 1 - \frac{\rho_j^{(m)}}{n} \right| \leqslant 2 \left| 1 - \frac{m r_j}{n} \right| + \frac{2m}{n} \left| \frac{\rho_j^{(m)}}{m} - r_j \right|.\end{aligned}$$

同时, 因为 $\rho_j^{(m)} \geqslant m$, 故有

$$\left| \overline{T}_j^{(\rho_j^{(m)})} - r_j^{-1} \right| \leqslant \frac{1}{r_j} \left| \frac{\rho_j^{(m)}}{m} - r_j \right|.$$

因此

$$\left| \overline{T}_j^{(n)} - r_j^{-1} \right| \leqslant 2 \left| 1 - \frac{m r_j}{n} \right| + \left(\frac{2m}{n} + \frac{1}{r_j} \right) \left| \frac{\rho_j^{(m)}}{m} - r_j \right|.$$

最后, 令 $m_n = \left[\frac{n}{r_j} \right]$, 可得当 $n \to \infty$ 时,

$$\left| \overline{T}_j^{(n)} - r_j^{-1} \right| \leqslant \frac{2}{n} + \frac{3}{r_j} \left| \frac{\rho_j^{(m_n)}}{m} - r_j \right| \to 0.$$

我们指出 (2.3.10) 式正是 Gibbs 所寻求的结果: 当观察一条路径时, 以概率 1, 它在每个状态上停留的平均时间随着观察的时间越来越长趋向于均衡 (平稳) 分布赋予那个状态的概率.

2.4 习题

2.4.1. 该习题将给出转移概率矩阵的*伴随矩阵*关于平稳分布的概率解释. 假定转移概率矩阵 $\boldsymbol{P}$ 有平稳分布 $\boldsymbol{\mu}$, 并假设对每个 $i \in \mathbb{S}$, $(\boldsymbol{\mu})_i > 0$, 由 $(\boldsymbol{P}^\top)_{ij} = \dfrac{(\boldsymbol{\mu})_j}{(\boldsymbol{\mu})_i}(\boldsymbol{P})_{ji}$ 确定矩阵 $\boldsymbol{P}^\top$.

(a) 证明 $\boldsymbol{P}^\top$ 是转移概率矩阵, 并且同样具有平稳分布 $\boldsymbol{\mu}$.

(b) 令 P 和 $\mathrm{P}^\top$ 分别表示具有初始分布 $\boldsymbol{\mu}$、转移概率矩阵 $\boldsymbol{P}$ 和 $\boldsymbol{P}^\top$ 的 Markov 链的概率. 证明这两个 Markov 链在下列意义下是*互逆的*: 对每个 $n \geqslant 0$, $(X_0, \cdots, X_n)$ 在 $\mathrm{P}^\top$ 下的概率分布与 $(X_n, \cdots, X_0)$ 在 P 下的概率分布相同. 换句话说, 对所有的 $n \geqslant 0$ 和 $(i_0, \cdots, i_n) \in \mathbb{S}^{n+1}$,

$$\mathrm{P}^\top(X_0 = i_0, \cdots, X_n = i_n) = \mathrm{P}(X_n = i_0, \cdots, X_0 = i_n).$$

2.4.2. Doeblin 理论特别适用于有限状态 Markov 链. 例如, 假定 $\boldsymbol{P}$ 是具有 N 个元素的状态空间 $\mathbb{S}$ 上的转移概率矩阵, 证明存在 $\epsilon > 0$ 使得对所有 $i \in \mathbb{S}$, $(\boldsymbol{A}_N)_{ij_0} \geqslant \epsilon$ 当且仅当对所有的 $i \in \mathbb{S}$, $i \to j_0$. 特别地, 如果这样的 j_0 存在, 那么对任意概率向量 $\boldsymbol{\mu}$,

$$\|\boldsymbol{\mu}\boldsymbol{A}_n - \boldsymbol{\pi}\|_v \leqslant \frac{2(N-1)}{n\epsilon}, \quad n \geqslant 1,$$

其中 $\boldsymbol{\pi}$ 是关于 $\boldsymbol{P}$ 的唯一平稳分布.

2.4.3. 下列形式的 Doeblin 定理有时候给出一个稍微好一点的估计. 假定对于任意的 (i, j) 有 $(\boldsymbol{P})_{ij} \geqslant \epsilon_j$, 令 $\epsilon = \sum\limits_j \epsilon_j$. 如果 $\epsilon > 0$, 证明定理 2.2.1 的结论成立, 且对每个 $i \in \mathbb{S}$ 有 $(\boldsymbol{\pi})_i \geqslant \epsilon_i$.

2.4.4. 假设 $\boldsymbol{P}$ 是有限状态空间 $\mathbb{S}$ 上的概率转移矩阵. 证明: $j \in \mathbb{S}$ 是常返的当且仅当 $\mathrm{E}[\rho_j | X_0 = j] < \infty$. 当然, 其中的充分性部分是显然的, 且与状态空间的有限性无关.

2.4.5. 假设 $\boldsymbol{P}$ 是有限状态空间 $\mathbb{S}$ 上的转移概率矩阵. 此外, 假设 $\boldsymbol{P}$ 是*双随机的*, 即它的每行、每列元素之和为 1. 证明, 如果每个状态都可以从其他状态到达, 那么对每个 $j \in \mathbb{S}$, $\mathrm{E}[\rho_j | X_0 = j] = \#\mathbb{S}$.

2.4.6. 为了检验 Doeblin 定理的效果, 考虑下列情形: $\mathbb{S}=\{1,2\}$, 对某个 $(\alpha,\beta)\in(0,1)$,

$$\boldsymbol{P}=\begin{pmatrix}1-\alpha & \alpha\\ \beta & 1-\beta\end{pmatrix}.$$

证明: $\boldsymbol{\pi}=(\alpha+\beta)^{-1}(\beta,\alpha)$,

$$\max\{\|\boldsymbol{\nu P}-\boldsymbol{\pi}\|_v : \boldsymbol{\nu}\ \text{是一个概率向量}\}=\frac{2(\alpha\vee\beta)|\alpha+\beta-1|}{\alpha+\beta}.$$

2.4.7. Markov 过程最早的一个例子是由 Galton 和 Watson 于 19 世纪末引入的分支过程, 用于建立人口模型. 在这个模型中,$\mathbb{S}=\mathbb{N}$, 状态 $i\in\mathbb{N}$ 表示人口的总数, 过程演变如下: 在每个时刻, 每个个体独立于其他个体死亡并由随机个数的下一代取代. 这样, 0 是一个吸收态, 并且如果在 n 时刻有 $i\geqslant 1$ 个个体活着, 到 $n+1$ 时刻的个体数与 i 个独立同分布、$\mathbb{N}$ 值的随机变量之和再减去 i 同分布. 更精确地, 设 $\boldsymbol{\mu}=(\mu_0,\cdots,\mu_k,\cdots)$ 是每个个体产生的下一代数目的概率向量, 定义 m 重卷积 $\boldsymbol{\mu}^{*m}$ 使得 $(\boldsymbol{\mu}^{*0})_j=\delta_{0,j}$, 并且对 $m\geqslant 1$,

$$(\boldsymbol{\mu}^{*m})_j=\sum_{i=0}^{j}(\boldsymbol{\mu}^{*(m-1)})_{j-i}\mu_i.$$

那么, 概率转移矩阵 $\boldsymbol{P}$ 由 $(\boldsymbol{P})_{ij}=(\boldsymbol{\mu}^{*i})_j$ 确定.

关于这个模型的第一个有趣的问题是, 最终灭绝的概率有多大? 也就是说, $\lim\limits_{n\to\infty}\mathrm{P}(X_n=0)$ 是什么? 一个自然的猜想是, 最终是否灭绝应该取决于期望值 $\gamma\equiv\sum\limits_{k=0}^{\infty}k\mu_k$ 是否严格小于 1 或者严格大于 1. 当 $\gamma=1$ 时, 情况更加复杂. 为了证实这个猜想, 并去掉一些明显的特殊情形, 我们假设 $(\boldsymbol{\mu})_0>0,(\boldsymbol{\mu})_0+(\boldsymbol{\mu})_1<1$, 且 $\gamma\equiv\sum\limits_{k=0}^{\infty}k(\boldsymbol{\mu})_k<\infty$.

(a) 对 $s\in[0,1]$, 令 $f(s)=\sum\limits_{k=0}^{\infty}s^k\mu_k$, 并且归纳地定义 $f^{\circ n}(s)$ 使得 $f^{\circ 0}(s)=s$, 对 $n\geqslant 1$, $f^{\circ n}=f\circ f^{\circ(n-1)}$. 证明: $\gamma=f'(1)$ 和

$$f^{\circ n}(s)^i=\mathrm{E}\left[s^{X_n}|X_0=i\right]=\sum_{j=0}^{\infty}s^j(\boldsymbol{P}^n)_{ij},\quad s\in[0,1],i\geqslant 0.$$

提示: 先证明 $f(s)^i=\sum\limits_{j=0}^{\infty}s^j(\boldsymbol{\mu}^{*i})_j$.

(b) 注意到 $s \in [0,1] \longmapsto f(s) - s$ 是一个连续函数, 在 $s = 0$ 处取正值, $s = 1$ 处等于 0, 它是光滑的并且在 $(0,1)$ 上严格凸 (即 $f'' > 0$). 证明: 或者 $\gamma \leqslant 1$ 并且对 $s \in [0,1)$, $f(s) > s$; 或者 $\gamma > 1$ 并且存在唯一的 $\alpha \in (0,1)$, 使得 $f(\alpha) = \alpha$.

(c) 继续上面, 证明:

$$\gamma \leqslant 1 \implies \lim_{n\to\infty} \mathrm{E}[s^{X_n}|X_0 = i] = 1, \quad s \in (0,1]$$

和

$$\gamma > 1 \implies \lim_{n\to\infty} \mathrm{E}[s^{X_n}|X_0 = i] = \alpha^i, \quad s \in (0,1).$$

(d) 基于 (c), 证明 $\gamma \leqslant 1 \implies \mathrm{P}(X_n = 0|X_0 = i) \to 1$ 和 $\gamma > 1 \implies \lim_{n\to\infty} \mathrm{P}(X_n = 0|X_0 = i) = \alpha^i$及

$$\lim_{n\to\infty} \mathrm{P}(1 \leqslant X_n \leqslant L|X_0 = i) = 0, \quad L \geqslant 1.$$

最后一个结论意味着, 当后代的期望值大于 1 时, 总体要么灭绝, 要么 (也许更糟糕) 无限膨胀.

2.4.8. 沿用习题 2.4.7. 的定义和记号. 我们想通过该习题说明,$\gamma < 1$ 和 $\gamma = 1$ 两者之间存在着显著差异.

(a) 证明 $\mathrm{E}[X_n|X_0 = i] = i\gamma^n$. 因此, 当 $\gamma < 1$ 时, 总体平均数以指数速度趋于 0. 另一方面, 当 $\gamma = 1$ 时, 尽管 $\mathrm{P}(X_n = 0|X_0 = i) \to 1$, 但总体平均数保持为常数. 这样, $\gamma = 1$ 提供了一个典型的情形, 说明 Lebesgue 为什么必须在他的控制收敛定理 (定理 6.1.11) 中作出假设. 在现在的例子中解释很简单: 当 $n \to \infty$ 时, $X_n = 0$ 的概率很大, 但是 X_n 以正概率趋于无穷.

(b) 令 ρ_0 为首次返回 0 的时刻. 证明

$$\mathrm{P}(\rho_0 \leqslant n|X_0 = i) = \mathrm{P}(X_n = 0|X_0 = i) = \left(f^{\circ(n-1)}(\mu_0)\right)^i,$$

并用它来得到估计

$$\mathrm{P}(\rho_0 > n|X_0 = i) \leqslant i\gamma^{n-1}(1 - \mu_0).$$

特别地, 这表明当 $\gamma < 1$, $\mathrm{E}[\rho_0|X_0 = i] < \infty$.

(c) 假设 $\gamma = 1$. 此外, 假设 $\beta \equiv f''(1) = \sum_{k\geqslant 2} k(k = 1)\mu_k < \infty$, 从 $\mathrm{P}(\rho_0 \leqslant n|X_0 = 1) = f^{\circ(n-1)}(\mu_0)$ 出发, 证明对所有 $i \geqslant 1$, $\mathrm{E}[\rho_0|X_0 =$

$i] = \infty$.
提示: 从证明对 $n > m$,

$$1 - f^{\circ n}(\mu_0) \geqslant \left(\prod_{\ell=m}^{n-1}\left(1 - \beta(1 - f^{\circ \ell}(\mu_0))\right)\right)(1 - f^{\circ m}(\mu_0))$$

开始. 然后利用它证明

$$\infty > \mathrm{E}[\rho_0 | X_0 = 1] = 1 + \sum_0^{\infty}(1 - f^{\circ n}(\mu_0)),$$

这将导致矛盾.

(**d**) 我们要证明: 一般情况下, 当去掉二阶导数有限的条件时, (**c**) 中的结论不成立. 给定 $\theta \in (0,1)$, 验证 $f(s) \equiv s + \dfrac{(1-s)^{1+\theta}}{1+\theta} = \sum\limits_{k=0}^{\infty} s^k \mu_k$, 其中 $\boldsymbol{\mu} = (\mu_0, \cdots, \mu_k, \cdots)$ 是概率向量, 并且除非 $k = 1$, $\mu_k > 0$. 现在根据 $\boldsymbol{\mu}$ 的选择可以看出, 当 (**c**) 中二阶导数条件不满足时, 即使 $\gamma = 1$, $\mathrm{E}[\rho_0 | X_0 = 1]$ 可能是有限的.
提示: 令 $a_n = 1 - f^{\circ n}(\mu_0)$, 注意到 $a_n - a_{n+1} = \mu_0 a_n^{1+\theta}$. 利用它首先证明 $\dfrac{a_{n+1}}{a_n} \to 1$, 再证明存在 $0 < c_1 < c_2 < \infty$ 使得对所有 $n \geqslant 1$, $c_1 \leqslant a_{n+1}^{-\theta} - a_n^{-\theta} \leqslant c_2$. 最后, 证明 $\mathrm{P}(\rho_0 > n | X_0 = 1)$ 以 $n^{-\frac{1}{\theta}}$ 的速度趋于 0.

2.4.9. 该习题中的思想由 Doob 引入, 并被称为 *Doob h-变换*.[7] 令 $\boldsymbol{P}$ 是状态空间 $\mathbb{S}$ 上的转移概率矩阵. 给定 $\varnothing \neq \Gamma \subset (\neq)\mathbb{S}$, 令 $\rho_\Gamma = \inf\{n \geqslant 1 : X_n \in \Gamma\}$, 并假设对所有 $i \in \hat{\mathbb{S}} \equiv \mathbb{S} \setminus \Gamma$,

$$h(i) \equiv \mathrm{P}(\rho_\Gamma = \infty | X_0 = i) > 0.$$

(**a**) 证明对所有 $i \in \hat{\mathbb{S}}$, $h(i) = \sum\limits_{j \in \hat{\mathbb{S}}} (\boldsymbol{P})_{ij} h(j)$, 并证明矩阵 $\hat{\boldsymbol{P}}$ 是 $\hat{\mathbb{S}}$ 上的转移概率矩阵, 其中 $(\hat{\boldsymbol{P}})_{ij} = \dfrac{1}{h(i)}(\boldsymbol{P})_{ij} h(j), \quad (i,j) \in (\hat{\mathbb{S}})^2$.

(**b**) 对所有 $n \in \mathbb{N}$ 和 $(j_0, \cdots, j_n) \in (\hat{\mathbb{S}})^{n+1}$, 证明: 对每个 $i \in \hat{\mathbb{S}}$, 成立

$$\begin{aligned}&\hat{\mathrm{P}}(X_0 = j_0, \cdots, X_n = j_n | X_0 = i) \\ &= \mathrm{P}(X_0 = j_0, \cdots, X_n = j_n | \rho_\Gamma = \infty, X_0 = i),\end{aligned}$$

[7] "h" 表示与调和函数有关.

其中 $\hat{\mathrm{P}}$ 表示 $\hat{\mathbb{S}}$ 上具有转移概率矩阵 $\hat{\boldsymbol{P}}$ 的 Markov 链的概率. 也就是说, 由 $\hat{\boldsymbol{P}}$ 决定的 Markov 链恰好是在永远不到达 Γ 的条件下, 由 $\boldsymbol{P}$ 决定的 Markov 链.

2.4.10. 下面是关于 h-变换的另一个例子. 假设 $j_0 \in \mathbb{S}$ 是瞬时的, 但是对所有 $i \in \mathbb{S}$, $i \to j_0$.[8] 对 $i \neq j_0$, 令 $h(i) = \mathrm{P}(\rho_{j_0} < \infty | X_0 = i)$, 并且 $h(j_0) = 1$.

(a) 验证对所有 $i \in \mathbb{S}$, $h(i) > 0$. 然后定义 $\hat{\boldsymbol{P}}$ 满足

$$(\hat{\boldsymbol{P}})_{ij} = \begin{cases} (\boldsymbol{P})_{j_0 j}, & \text{若 } i = j_0; \\ h(i)^{-1}(\boldsymbol{P})_{ij} h(j), & \text{若 } i \neq j_0. \end{cases}$$

证明 $\hat{\boldsymbol{P}}$ 是转移概率矩阵.

(b) 用 $\hat{\mathrm{P}}$ 表示由 $\hat{\boldsymbol{P}}$ 决定的 Markov 链的概率, 证明对任意 $n \in \mathbb{N}$ 及 $i \neq j_0$, 成立

$$\hat{\mathrm{P}}(\rho_{j_0} > n | X_0 = i) = \frac{1}{h(i)} \mathrm{P}(n < \rho_{j_0} < \infty | X_0 = i).$$

(c) 从 (b) 中的结果出发, 证明对于由 $\hat{\boldsymbol{P}}$ 决定的 Markov 链, j_0 是常返状态.

[8] 由习题 2.4.2 知, 仅当 $\mathbb{S}$ 是无穷状态空间时, 这才可能.

第三章　Markov 链的遍历理论(续)

在第二章, 所有的讨论都是围绕 Doeblin 条件展开的, 该条件意味着存在一个状态, 可以从任何一个其他状态以一致快的速度到达它. 尽管有许多无穷状态空间上的 Markov 链满足他的条件, 但大多数并不满足. 因此, 许多无穷状态空间上的 Markov 链不具有平稳概率分布. 事实上, 具有无穷多状态意味着该链有足够的空间 "迷失和消失". 这可以两种方式发生. 就是说, Markov 链会消失, 因为它没有常返状态, 从而导致在任意给定的状态上只停留有限的时间, 比如像 $\mathbb{Z}$ 上的最近邻、非对称的随机游动 (参看 (1.1.13)) 或者甚至 $\mathbb{Z}^3$ 上的对称随机游动 (参看 §1.2.4). Markov 链会消失的另一种更加微妙的方式是, 它有常返状态但是却不足以使该链具有平稳概率分布. 这样的一个例子是 $\mathbb{Z}$ 上的最近邻对称随机游动, 它是常返的, 但仅此而已. 特别地, 尽管该随机游动可以无穷多次返回它的初始点, 但是它在一个非常稀疏的时刻集合上返回. 更精确地, 根据 (1.2.7) 式和 (1.2.13) 式, 如果 $\boldsymbol{P}$ 是 $\mathbb{Z}$ 上的最近邻对称随机游动的转移概率矩阵, 那么

$$(\boldsymbol{P}^{2n})_{ij} \leqslant (\boldsymbol{P}^{2n})_{ii} = \mathrm{P}(X_{2n} = 0) \leqslant A(1)n^{-\frac{1}{2}} \to 0.$$

因此如果 $\boldsymbol{\mu}$ 是概率向量且关于 $\boldsymbol{P}$ 是平稳的, 那么由 Lebesgue 控制收

敛定理 (定理 6.1.11), 产生矛盾

$$(\boldsymbol{\mu})_j = \lim_{n\to\infty} \sum_{i\in\mathbb{Z}} (\boldsymbol{\mu})_i (\boldsymbol{P}^{2n})_{ij} = 0, \quad \forall j \in \mathbb{Z}.$$

在这一章里, 我们将看到即使不假设 Doeblin 条件, 也可有遍历性. 然而, 正如我们将看到的那样, 它以一种较弱的形式成立. 特别地, 将不再寻求依范数 $\|\cdot\|_v$ 的收敛性, 而代之以逐点收敛性. 也就是说, 将寻求对每个 j, $(\boldsymbol{\mu P})_j \to (\boldsymbol{\pi})_j$ 的结果, 而不是 $\|\boldsymbol{\mu P} - \boldsymbol{\pi}\|_v \to 0$.

3.1 状态的分类

在这一节, 我们将处理在第二章中已经提示但没有明确讨论的问题. 因为不再假设 Doeblin 条件, 所以有必要考虑这样的可能性, 即把 Markov 链的状态空间看成若干个互不相交子集的并. 确切地说, 给定一对状态 (i, j), 如果 Markov 链能够以正概率从状态 i 到达状态 j, 那么记 $i \to j$ 并称 j 可从 i 到达. 即对某个 $n \in \mathbb{N}$, $(\boldsymbol{P}^n)_{ij} > 0$. 注意, 可到达性具有传递性:

$$i \to j, j \to \ell \implies i \to \ell. \tag{3.1.1}$$

事实上, 如果 $(\boldsymbol{P}^m)_{ij} > 0$ 且 $(\boldsymbol{P}^n)_{j\ell} > 0$, 则

$$(\boldsymbol{P}^{m+n})_{i\ell} = \sum_k (\boldsymbol{P}^m)_{ik} (\boldsymbol{P}^n)_{k\ell} \geqslant (\boldsymbol{P}^m)_{ij} (\boldsymbol{P}^n)_{j\ell} > 0.$$

如果 i 和 j 是彼此可到达的, 即 $i \to j$ 和 $j \to i$, 那么我们记 $i \leftrightarrow j$ 并称 i 与 j *互通*. 很明显, $\leftrightarrow$ 是一个等价关系. 事实上, 因为 $(\boldsymbol{P}^0)_{ii} = 1$, 所以 $i \leftrightarrow i$; 另外, 如果 $i \leftrightarrow j$, 那么 $j \leftrightarrow i$. 最后, 如果 $i \leftrightarrow j$ 和 $j \leftrightarrow \ell$, 则由 (3.1.1) 式明显得到 $i \leftrightarrow \ell$. 因此, $\leftrightarrow$ 把整个状态空间分成了互通的等价类. 也就是说, 对任意状态 i, 互通等价类 $[i]$ 是由所有和 i 互通的状态 j 构成的集合, 并且对任意一对状态 (i, j), $[i] = [j]$ 或者 $[i] \bigcap [j] = \varnothing$. 当每个状态都和其他状态互通时, 称该链是*不可约的*.

3.1.1. 分类、常返性和瞬时性: 在这一小节, 将证明常返性和瞬时性是*互通类性质*. 也就是说, 一个互通等价类中的所有状态要么都是常返的, 要么都是瞬时的.

回忆一下 (参看 §2.3.2), ρ_j 是首次返回到 j 的时刻, 并根据 $\mathbb{P}(\rho_j < \infty | X_0 = j)$ 等于 1 或严格小于 1, 称 j 是*常返的*或*瞬时的*.

3.1.2 定理. 假设 i 是常返的, $j \neq i$. 则 $i \to j$ 当且仅当 $\mathrm{P}(\rho_j < \rho_i|X_0 = i) > 0$. 进一步, 如果 $i \to j$, 那么对任意 $(k,\ell) \in \{i,j\}^2$,

$$\mathrm{P}(\rho_k < \infty|X_0 = \ell) = 1.$$

特别地, $i \to j$ 可以推出 $i \leftrightarrow j$ 并且 j 是常返的.

证明: 给定 $j \neq i, n \geqslant 1$, 令 (参看 (2.3.5))

$$G_n(k_0,\cdots,k_n) = (F_{n-1,i}(k_0,\cdots,k_{n-1}) - F_{n,i}(k_0,\cdots,k_n)) \cdot F_{n,j}(k_0,\cdots,k_n).$$

如 §2.3.2 中那样定义 $\{\rho_j^{(m)} : m \geqslant 0\}$, 那么由 (2.1.2) 式得

$$\begin{aligned}
&\mathrm{P}(\rho_i^{(m+1)} < \rho_j|X_0 = i)\\
&= \sum_{\ell=1}^{\infty} \mathrm{P}(\rho_i^{(m)} = \ell, \rho_i^{(m+1)} < \rho_j|X_0 = i)\\
&= \sum_{\ell=1}^{\infty}\sum_{n=1}^{\infty} \mathrm{P}(\rho_i^{(m)} = \ell, \rho_i^{(m+1)} = \ell + n < \rho_j|X_0 = i)\\
&= \sum_{\ell,n=1}^{\infty} \mathrm{E}\left[G_n(X_\ell,\cdots,X_{\ell+n}), \rho_i^{(m)} = \ell < \rho_j|X_0 = i\right]\\
&= \sum_{\ell,n=1}^{\infty} \mathrm{E}\left[G_n(X_0,\cdots,X_n)|X_0 = i\right] \mathrm{P}(\rho_i^{(m)} = \ell < \rho_j|X_0 = i)\\
&= \sum_{\ell,n=1}^{\infty} \mathrm{P}(\rho_i = n < \rho_j|X_0 = i)\mathrm{P}(\rho_i^{(m)} = \ell < \rho_j|X_0 = i)\\
&= \mathrm{P}(\rho_i < \rho_j|X_0 = i)\mathrm{P}(\rho_i^{(m)} < \rho_j|X_0 = i),
\end{aligned}$$

因此

$$j \neq i \implies \mathrm{P}(\rho_i^{(m)} < \rho_j|X_0 = i) = \mathrm{P}(\rho_i < \rho_j|X_0 = i)^m. \tag{3.1.3}$$

假设 $i \to j$ 但 $\mathrm{P}(\rho_j < \rho_i|X_0 = i) = 0$. 那么 $\mathrm{P}(\rho_i \leqslant \rho_j|X_0 = i) = 1$. 由此并注意到 $\mathrm{P}(\rho_j \neq \rho_i|X_0 = i) \geqslant \mathrm{P}(\rho_i < \infty|X_0 = i) = 1$, 所以 $\mathrm{P}(\rho_i < \rho_j|X_0 = i) = 1$. 因此, 由 (3.1.3) 式得, 对所有 $m \geqslant 1$, $\mathrm{P}(\rho_i^{(m)} < \rho_j|X_0 = i) = 1$. 回顾 $\rho^{(m)} \geqslant m$, 进而得到 $\mathrm{P}(\rho_j = \infty|X_0 = i) = 1$, 从而 $i \to j$ 是不成立的. 至此, 已经证明了 $i \to j \implies \mathrm{P}(\rho_j < \rho_i|X_0 = i) > 0$. 相反方向的关系显然成立.

为了证明 $i \to j \implies \mathrm{P}(\rho_i < \infty | X_0 = j) = 1$, 首先注意到

$$
\begin{aligned}
&\mathrm{P}(\rho_j < \rho_i < \infty | X_0 = i) \\
&= \lim_{n\to\infty} \sum_{m=1}^{\infty} \mathrm{P}(\rho_j = m < \rho_i \leqslant m + n | X_0 = i) \\
&= \lim_{n\to\infty} \sum_{m=1}^{\infty} \mathrm{E}\left[1 - F_{n,i}(X_m, \cdots, X_{m+n}), \rho_j = m < \rho_i | X_0 = i\right] \\
&= \lim_{n\to\infty} \sum_{m=1}^{\infty} \mathrm{P}(\rho_j = m < \rho_i | X_0 = i)\mathrm{E}\left[1 - F_{n,i}(X_0, \cdots, X_n) | X_0 = j\right] \\
&= \mathrm{P}(\rho_j < \rho_i | X_0 = i)\mathrm{P}(\rho_i < \infty | X_0 = j).
\end{aligned}
$$

这样, 结合 $\mathrm{P}(\rho_i < \infty | X_0 = i) = 1$, 我们有

$$
\mathrm{P}(\rho_j < \rho_i | X_0 = i) = \mathrm{P}(\rho_j < \rho_i | X_0 = i)\mathrm{P}(\rho_i < \infty | X_0 = j).
$$

因为 $\mathrm{P}(\rho_j < \rho_i | X_0 = i) > 0$, 所以上式仅当 $\mathrm{P}(\rho_i < \infty | X_0 = j) = 1$ 时才能成立. 特别地, 我们证明了 $j \to i$, 所以 $i \leftrightarrow j$.

类似地,

$$
\begin{aligned}
&\mathrm{P}(\rho_j < \infty | X_0 = i) \\
&= \mathrm{P}(\rho_j < \rho_i | X_0 = i) + \mathrm{P}(\rho_i < \rho_j < \infty | X_0 = i) \\
&= \mathrm{P}(\rho_j < \rho_i | X_0 = i) + \mathrm{P}(\rho_i < \rho_j | X_0 = i)\mathrm{P}(\rho_j < \infty | X_0 = i),
\end{aligned}
$$

因此

$$
\mathrm{P}(\rho_j < \infty | X_0 = i)\mathrm{P}(\rho_j < \rho_i | X_0 = i) = \mathrm{P}(\rho_j < \rho_i | X_0 = i).
$$

这样, $i \to j \implies \mathrm{P}(\rho_j < \infty | X_0 = i) = 1$.

最后,

$$
\mathrm{P}(\rho_i < \rho_j < \infty | X_0 = j) = \mathrm{P}(\rho_j < \infty | X_0 = i)\mathrm{P}(\rho_i < \rho_j | X_0 = j).
$$

因此, 既然当 $i \to j$ 时, $\mathrm{P}(\rho_i < \infty | X_0 = j) = 1 = \mathrm{P}(\rho_j < \infty | X_0 = i)$, 那么由于 $\mathrm{P}(\rho_i = \rho_j | X_0 = j) \leqslant \mathrm{P}(\rho_i = \infty | X_0 = j) = 0$, 可以看出 $i \to j$ 蕴涵了

$$
\begin{aligned}
&\mathrm{P}(\rho_j < \infty | X_0 = j) \\
&= \mathrm{P}(\rho_j < \rho_i | X_0 = j) + \mathrm{P}(\rho_i < \rho_j < \infty | X_0 = j) \\
&= \mathrm{P}(\rho_j < \rho_i | X_0 = j) + \mathrm{P}(\rho_i < \infty | X_0 = i)\mathrm{P}(\rho_i < \rho_j | X_0 = j) \\
&= 1. \quad \square
\end{aligned}
$$

作为定理 3.1.2 的直接结果, 我们有下面推论.

3.1.4 推论. 如果 $i \leftrightarrow j$, 则 j 是常返的 (或者瞬时的) 当且仅当 i 是常返的 (或者瞬时的). 此外, 如果 i 是常返的, 则当 i 和 j 互通时,$\mathrm{P}(\rho_j < \infty | X_0 = i)$ 等于 1, 否则等于 0. 特别地, 如果 i 是常返的, 则对所有 $n \geqslant 0$ 和与 i 不互通的状态 j, 有 $(\boldsymbol{P}^n)_{ij} = 0$.

当 Markov 链不可约时, 所有状态都具有给定的互通类性质, 或者一个也不具有给定的互通类性质. 因此, 当 Markov 链不可约时, 如果其中一个状态 (因此所有状态) 是常返的或者瞬时的, 则称整个 Markov 链是常返的 或者瞬时的.

3.1.2. 常返性和瞬时性的判别法则: 有许多检验方法可以帮助确定一个状态是否常返, 但是没有一种方法能够适用于所有情况. 在这一小节, 我们将给出其中最常用的一些判别方法. 在下文中, 令 $\boldsymbol{u}$ 表示由函数 $u : \mathbb{S} \to \mathbb{R}$ 所确定的列向量.

首先给出一个瞬时性的判别法则.

3.1.5 定理. 设 $\mathbb{S}$ 上的非负函数 u 满足: 对所有 $i \in \mathbb{S}$, $(\boldsymbol{Pu})_i \leqslant (\boldsymbol{u})_i$. 那么对某一个 $j \in \mathbb{S}$, $(\boldsymbol{Pu})_j < (\boldsymbol{u})_j$ 推出 j 是瞬时的.

证明: 令 $\boldsymbol{f} = \boldsymbol{u} - \boldsymbol{Pu}$. 注意到对所有 $n \geqslant 1$,

$$
\begin{aligned}
u(j) \geqslant (\boldsymbol{u})_j - (\boldsymbol{P}^n\boldsymbol{u})_j &= \sum_{m=0}^{n-1}((\boldsymbol{P}^m\boldsymbol{u})_j - (\boldsymbol{P}^{m+1}\boldsymbol{u})_j) \\
&= \sum_{m=0}^{n-1}(\boldsymbol{P}^m\boldsymbol{f})_j \geqslant \sum_{m=0}^{n-1}(\boldsymbol{P}^m)_{jj},
\end{aligned}
$$

因此 $\mathrm{E}[T_j|X_0 = j] = \sum\limits_{m=0}^{\infty}(\boldsymbol{P}^m)_{jj} \leqslant \dfrac{u(j)}{(\boldsymbol{f})_j} < \infty$. 由 (2.3.7) 式知, j 是瞬时的. □

为了证明下一个判别法则, 我们需要如下 *Doob* 停时定理的一个特殊情形.

3.1.6 引理. 假设 $u : \mathbb{S} \to \mathbb{R}$ 是下有界的, Γ 是 $\mathbb{S}$ 的非空子集. 如果对所有 $i \notin \Gamma$, $(\boldsymbol{Pu})_i \leqslant u(i)$, $\rho_\Gamma \equiv \inf\{n \geqslant 1 : X_n \in \Gamma\}$, 那么对所有 $n \geqslant 0$ 和 $i \in \mathbb{S} \setminus \Gamma$,

$$
\mathrm{E}[u(X_{n\wedge\rho_\Gamma})|X_0 = i] \leqslant u(i).
$$

此外, 如果假设中的不等号改成等号, 则结论中的不等号变成等号.

证明: 令 $A_n = \{\rho_\Gamma > n\}$. 那么 A_n 关于 $(X_0, \cdots, X_n)$ 是可测的. 由

(2.1.1) 式知, 对任意的 $i \notin \Gamma$,

$$
\begin{aligned}
&\mathrm{E}[u(X_{(n+1)\wedge\rho_\Gamma})|X_0=i] \\
&=\mathrm{E}[u(X_{n\wedge\rho_\Gamma}),A_n^c|X_0=i]+\sum_{k\notin\Gamma}\mathrm{E}[u(X_{n+1}),A_n\bigcap\{X_n=k\}|X_0=i] \\
&=\mathrm{E}[u(X_{n\wedge\rho_\Gamma}),A_n^c|X_0=i]+\sum_{k\notin\Gamma}\mathrm{E}[(\boldsymbol{Pu})_k,A_n\bigcap\{X_n=k\}|X_0=i] \\
&\leqslant \mathrm{E}[u(X_{n\wedge\rho_\Gamma}),A_n^c|X_0=i]+\mathrm{E}[u(X_{n\wedge\rho_\Gamma}),A_n|X_0=i] \\
&=\mathrm{E}[u(X_{n\wedge\rho_\Gamma})|X_0=i].
\end{aligned}
$$

显然, 对于等式的情形, 上面的论证同样成立. □

3.1.7 定理. *假设 j 是常返的, 令 $C=\{i:i\leftrightarrow j\}$. 如果 $u:\mathbb{S}\to[0,\infty)$ 是有界函数, 并且对任意 $i\in C\setminus\{j\}$, $u(i)=(\boldsymbol{Pu})_i$ 或者 $u(j)\geqslant u(i)\geqslant(\boldsymbol{Pu})_i$, 则 u 在 C 上为常数. 另一方面, 如果 j 是瞬时的, 那么对任意 $i\neq j$, 由*

$$
u(i)=\begin{cases}1, & 若\ i=j; \\ \mathrm{P}(\rho_j<\infty|X_0=i), & 若\ i\neq j\end{cases}
$$

确定的函数 u 是方程 $u(i)=(\boldsymbol{Pu})_i$ 的有界、非常数解.

证明: 为了证明第一部分, 不失一般性, 假设 $C=\mathbb{S}$. 设 j 是常返的, 并且对 $i\neq j$, $u(i)=(\boldsymbol{Pu})_i$. 令 $\Gamma=\{j\}$, 运用引理 3.1.6 得, 对 $i\neq j$,

$$
u(i)=u(j)\mathrm{P}(\rho_j\leqslant n|X_0=i)+\mathrm{E}[u(X_n),\rho_j>n|X_0=i].
$$

因此, 既然由定理 3.1.2, $\mathrm{P}(\rho_j<\infty|X_0=i)=1$, 并且 u 是有界的, 那么令 $n\to\infty$ 立即得到 $u(i)=u(j)$. 下面假设对所有 $i\neq j$, $u(j)\geqslant u(i)\geqslant(\boldsymbol{Pu})_i$. 那么再次运用引理 3.1.6 得

$$
u(j)\geqslant u(i)\geqslant u(j)\mathrm{P}(\rho_j\leqslant n|X_0=i)+\mathrm{E}[u(X_n),\rho_j>n|X_0=i].
$$

令 $n\to\infty$ 即可得到第一个结论.

为了证明第二个结论, 令 u 如上面的定义, 并注意到, 因为 j 是瞬时的, 所以

$$
1>\mathrm{P}(\rho_j<\infty|X_0=j)=\boldsymbol{P}_{jj}+\sum_{i\neq j}\boldsymbol{P}_{ji}u(i)\geqslant\boldsymbol{P}_{jj}+(1-\boldsymbol{P}_{jj})\inf_{i\neq j}u(i).
$$

由上式知, $\boldsymbol{P}_{jj}<1$ 且 $\inf_{i\neq j}u(i)<1=u(j)$. 也就是说, u 是有界、非常数的. 同时, 当 $i\neq j$ 时, 通过对 1 时刻所发生的事件进行条件概率运

算, 得

$$
\begin{aligned}
u(i) &= \mathrm{P}(\rho_j < \infty | X_0 = i) \\
&= \boldsymbol{P}_{ij} + \sum_{k \neq j} \boldsymbol{P}_{ik} \mathrm{P}(\rho_j < \infty | X_0 = k) = (\boldsymbol{P}\boldsymbol{u})_i. \quad \square
\end{aligned}
$$

3.1.8 引理. *若在 $\mathbb{S}$ 上 $\boldsymbol{P}$ 是不可约的, 则对任意的有限子集 $F \neq \mathbb{S}$, $\mathrm{P}(\rho_{\mathbb{S}\setminus F} < \infty | X_0 = i) = 1$ 对所有 $i \in F$ 成立.*

证明: 记 $\tau = \rho_{\mathbb{S}\setminus F}$. 由不可约性, 对每个 $i \in F$ 有 $\mathrm{P}(\tau < \infty | X_0 = i) > 0$. 因此, 由 F 的有限性知, 存在一个 $\theta \in (0,1)$ 和一个整数 $N \geqslant 1$ 使得 $\mathrm{P}(\tau > N | X_0 = i) \leqslant \theta$ 对所有 $i \in F$ 成立. 而这意味着, 对每个 $i \in F$, $\mathrm{P}(\tau > (\ell+1)N | X_0 = i)$ 等于

$$
\begin{aligned}
&\sum_{k \in F} \mathrm{P}(\tau > (\ell+1)N,\ X_{\ell N} = k | X_0 = i) \\
&= \sum_{k \in F} \mathrm{P}(\text{对 } \ell N + 1 \leqslant n \leqslant (\ell+1)N \text{ 有 } X_n \in F, \\
&\quad \tau > \ell N,\ X_{\ell N} = k | X_0 = i) \\
&= \sum_{k \in F} \mathrm{P}(\tau > N | X_0 = k) \mathrm{P}(\tau > \ell N,\ X_{\ell N} = k | X_0 = i) \\
&\leqslant \theta \mathrm{P}(\tau > \ell N | X_0 = i).
\end{aligned}
$$

于是, $\mathrm{P}(\tau > \ell N | X_0 = i) \leqslant \theta^\ell$, 从而对所有 $i \in F$ 有 $\mathrm{P}(\tau = \infty | X_0 = i) = 0$. $\square$

3.1.9 定理. *假设在 $\mathbb{S}$ 上 $\boldsymbol{P}$ 是不可约的. 令 $u: \mathbb{S} \to [0,\infty)$ 为一个函数, 具有如下性质: 对每个 $L \in (0,\infty)$, $\{k : u(k) \leqslant L\}$ 是有限的. 如果对某个 $j \in \mathbb{S}$, $(\boldsymbol{P}\boldsymbol{u})_i \leqslant u(i)$ 对所有 $i \neq j$ 成立, 那么由 $\boldsymbol{P}$ 确定的 Markov 链是常返的.*

证明: 如果 $\mathbb{S}$ 是有限的, 那么 (参见习题 2.4.2) 至少有一个状态是常返的. 从而, 由不可约性, 所有状态是常返的. 因此, 下面假设 $\mathbb{S}$ 是无限的.

给定 $i \neq j$, 对 $L \in \mathbb{N}$, 记 $F_L = \{k : u(k) \leqslant u(i) + u(j) + L\}$, 并用 ρ_L 表示首次返回 $(\mathbb{S} \setminus F_L) \bigcup \{j\}$ 的返回时 $\rho_{(\mathbb{S}\setminus F_L)\bigcup\{j\}}$. 由引理 3.1.6, 对所有的 $n \geqslant 1$ 有

$$
\begin{aligned}
u(i) &\geqslant \mathrm{E}\big[u(X_{n \wedge \rho_L}) \big| X_0 = i\big] \\
&\geqslant \big(u(i) + u(j) + L\big) \mathrm{P}\big(\rho_{\mathbb{S}\setminus F_L} < n \wedge \rho_j \big| X_0 = i\big).
\end{aligned}
$$

因为由引理 3.1.8 我们有 $\mathrm{P}(\rho_{\mathbb{S}\setminus F_L} < \infty | X_0 = i) = 1$, 所以, 令 $n \to \infty$ 得, 对所有的 $L \in \mathbb{N}$ 有

$$\begin{aligned} u(i) &\geqslant \big(u(i) + u(j) + L\big)\mathrm{P}\big(\rho_{\mathbb{S}\setminus F_L} < \rho_j \big| X_0 = i\big) \\ &\geqslant \big(u(i) + u(j) + L\big)\mathrm{P}\big(\rho_j = \infty \big| X_0 = i\big). \end{aligned}$$

于是, 我们得证: 对所有的 $i \neq j$ 有 $\mathrm{P}(\rho_j < \infty | X_0 = i) = 1$. 由于 $\mathrm{P}(\rho_j < \infty | X_0 = j) = (\boldsymbol{P})_{jj} + \sum_{i \neq j} \mathrm{P}(\rho_j < \infty | X_0 = i)(\boldsymbol{P})_{ji}$, 所以 $\mathrm{P}(\rho_j < \infty | X_0 = j) = 1$, 此即意味着 j 是常返的.

注: 上面的判别法则只是一些例子, 它们把 $j \in \mathbb{S}$ 的常返性和在 $\mathbb{S} \setminus \{j\}$ 上满足 $\boldsymbol{Pu} = \boldsymbol{u}$ 或者 $\boldsymbol{Pu} \leqslant \boldsymbol{u}$ 的某种类型函数的存在性或不存在性联系起来, 还有许多这类其他例子. 所有这些判别法则可以看做是下面直观想法的数学描述, 即

$$u(i) = (\boldsymbol{Pu})_i \text{ 或者 } (\boldsymbol{Pu})_i \leqslant u(i), \quad i \neq j,$$

蕴涵了只要 $X_n \neq j$, 那么随着 n 的递增, $u(X_n)$ "几乎不变" 或者 "几乎不增". 这种 "几乎" 的含义是鞅论中的内容, 如果能运用鞅论的知识, 这些法则的证明将会简化.

3.1.3. 周期性: 周期性是另一个重要的互通类性质. 为了叙述这个性质, 回顾一下 Euclid 关于非空子集 $S \subseteq \mathbb{Z}$ 的 **最大公约数** $\gcd(S)$ 的概念. 如果对所有 $s \in S$, $\dfrac{s}{d} \in \mathbb{Z}$, 我们说 $d \in \mathbb{Z}^+$ 是 S 的一个公约数, 并记为 $d|S$. 显然, 如果 $S = \{0\}$, 则对任意 $d \in \mathbb{Z}^+$, 有 $d|S$, 因此定义 $\gcd(S) = \infty$. 另一方面, 如果 $S \neq \{0\}$, 那么 S 的公约数中没有一个会超过 $\min\{|s| : s \in S \setminus \{0\}\}$, 从而得证 $\gcd(S) < \infty$.

我们对于这个概念的兴趣源于它在 Markov 链的遍历理论中所起的作用. 正如下面将看到的那样, 它有助于把转移概率矩阵幂收敛的 Markov 链和必须取平均后才收敛的 Markov 链区分开来. 更准确地, 给定状态 i, 令

$$S(i) = \{n \geqslant 0 : (\boldsymbol{P}^n)_{ii} > 0\} \text{ 及 } d(i) = \gcd(S(i)). \tag{3.1.10}$$

则 $d(i)$ 被称为状态 i 的**周期**; 如果 $d(i) = 1$, 称状态 i 是**非周期的**.

正如将看到的那样, 除非 i 是非周期的, 取平均是必要的. 然而, 在和遍历理论联系起来之前, 需要考虑一些平凡的性质. 首先, 周期是

一种互通类的性质:

$$i \leftrightarrow j \implies d(i) = d(j). \tag{3.1.11}$$

为了证明上式, 假设 $(\boldsymbol{P}^m)_{ij} > 0$ 和 $(\boldsymbol{P}^n)_{ji} > 0$, 令 d 是 $S(i)$ 的公约数. 那么对任意 $k \in S(j)$, $(\boldsymbol{P}^{m+k+n})_{ii} \geqslant (\boldsymbol{P}^m)_{ij}(\boldsymbol{P}^k)_{jj}(\boldsymbol{P}^n)_{ji} > 0$, 所以 $m+k+n \in S(i)$. 因此 $d|\{m+k+n : k \in S(j)\}$. 但是, 因为 $m+n \in S(i)$, 并由此 $d|(m+n)$, 因此仅当 $d|S(j)$ 时这才可能成立, 从而可得 $d(i) \leqslant d(j)$. 交换 i 和 j 后又可得 $d(j) \leqslant d(i)$. 这意味着 $d(i)$ 一定等于 $d(j)$.

我们需要下面数论中的初等事实.

3.1.12 定理. *给定 $\varnothing \neq S \subseteq \mathbb{Z}$ 且 $S \neq \{0\}$, 那么 $\gcd(S) \leqslant \min\{|s| : s \in S \setminus \{0\}\}$, 并且等式成立当且仅当 $\{\gcd(S), -\gcd(S)\} \bigcap S \neq \varnothing$. 更一般地, 总存在一个 $M \in \mathbb{Z}^+$, $\{a_m\}_1^M \subseteq \mathbb{Z}$ 和 $\{s_m\}_1^M \subseteq S$ 使得 $\gcd(S) = \sum\limits_1^M a_m s_m$. 最后, 如果 $S \subseteq \mathbb{N}$ 及 $(s_1, s_2) \in S^2 \implies s_1 + s_2 \in S$, 那么存在一个 $M \in \mathbb{Z}^+$ 使得*

$$\{m \gcd(S) : m \geqslant M\} = \{s \in S : s \geqslant M \gcd(S)\}.$$

证明: 第一个结论显然成立. 为了证明第二个结论, 令 $\hat{S}$ 是 $\mathbb{Z}$ 中包含 S 并且具有性质 $(s_1, s_2) \in \hat{S}^2 \implies s_1 \pm s_2 \in \hat{S}$ 的最小子集. 容易验证, $\hat{S}$ 与 $\mathbb{Z}$ 中具有如下形式的元素组成的子集相同: $\sum\limits_1^M a_m s_m$, 其中 $M \geqslant 1$, $\{a_m\}_1^M \subseteq \mathbb{Z}$, $\{s_m\}_1^M \subseteq S$. 特别地, 这意味着 $\gcd(S)|\hat{S}$, 并且由此得 $\gcd(S) \leqslant \gcd(\hat{S})$. 另一方面, 因为 $S \subseteq \hat{S}$, 所以 $\gcd(\hat{S})|S$. 因此 $\gcd(S) = \gcd(\hat{S})$, 并由第一个结论, 只要证明 $\gcd(\hat{S}) \in \hat{S}$ 就够了. 为此, 令 $m = \min\{s \in \mathbb{Z}^+ : s \in \hat{S}\}$. 已知 $\gcd(\hat{S}) \leqslant m$. 这样, 为了证明等号成立, 只要验证 $m|\hat{S}$. 但是, 由 Euclid 运算法则, 对任意 $s \in \hat{S}$, 可以写成 $s = am + r$, 其中某 $(a, r) \in \mathbb{Z}^2$, $0 \leqslant r \leqslant m$. 特别地, $r = s - am \in \hat{S}$. 因此, 如果 $r \neq 0$, 那么 r 与 m 是 $\hat{S}$ 的最小正元素这一条件矛盾.

为了证明最后一个结论, 只要证明存在一个 $M \in \mathbb{Z}^+$ 使得 $\{m \gcd(S) : m \geqslant M\} \subseteq S$ 就够了. 为此, 先验证在假设条件下, $\hat{S} = \{s_2 - s_1 : (s_1, s_2) \in S \bigcup \{0\}\}$. 这样, 对某个 $s_1 \in S \bigcup \{0\}$ 和 $s_2 \in S \setminus \{0\}$, $\gcd(S) = s_2 - s_1$. 如果 $s_1 = 0$, 则对任意 $m \in \mathbb{Z}^+$, 有 $m \gcd(S) = m s_2 \in S$. 因此可以取 $M = 1$. 如果 $s_1 \neq 0$, 选择 $a \in \mathbb{Z}^+$

使得 $s_1 = a\gcd(S)$. 从而, 对任意 $(m,r) \in \mathbb{N}^2$, 其中 $0 \leqslant r < a$, 有

$$\begin{aligned}(a^2 + ma + r)\gcd(S) &= ms_1 + rs_2 + (a-r)s_1 \\ &= (m + a - r)s_1 + rs_2 \in S.\end{aligned}$$

再次利用 Euclid 运算法则, 我们可以取 $M = a^2$. □

作为定理 3.1.12 的一个直接推论, 对所有充分大的 $n \in \mathbb{Z}^+$, 我们有

$$d(i) < \infty \implies (\boldsymbol{P}^{nd(i)})_{ii} > 0. \tag{3.1.13}$$

特别地[1],

$$i \text{ 是非周期的} \iff \text{对所有充分大的 } n \in \mathbb{Z}^+, \quad (\boldsymbol{P}^n)_{ii} > 0. \tag{3.1.14}$$

作为本小节的结尾, 我们将上述讨论应用到有限状态空间 Markov 链的遍历理论.

3.1.15 推论. *假设 $\boldsymbol{P}$ 是有限状态空间 $\mathbb{S}$ 上的转移概率矩阵. 如果存在非周期状态 $j_0 \in \mathbb{S}$, 使得对任意 $i \in \mathbb{S}$, 有 $i \to j_0$. 那么存在一个 $M \in \mathbb{Z}^+$ 和 $\epsilon > 0$ 使得对任意的 $i \in \mathbb{S}$, 有 $(\boldsymbol{P}^M)_{ij_0} \geqslant \epsilon$. 特别地 (参看 (2.3.9)), 对所有 $n \in \mathbb{Z}$ 和初始分布 $\boldsymbol{\mu}$,*

$$\|\boldsymbol{\mu}\boldsymbol{P}^n - \boldsymbol{\pi}\|_v \leqslant 2(1-\epsilon)^{[\frac{n}{M}]}.$$

证明: 因为 j_0 是非周期的, 所以存在 $M_0 \in \mathbb{N}$ 使得对任意 $n \geqslant M_0$, $(\boldsymbol{P}^n)_{j_0j_0} > 0$. 进而, 因为 $i \to j_0$, 所以存在 $m(i) \in \mathbb{Z}^+$ 使得 $(\boldsymbol{P}^{m(i)})_{ij_0} > 0$. 因此, 对任意的 $n \geqslant m(i) + M_0$ 有 $(\boldsymbol{P}^n)_{ij_0} > 0$. 最后, 取 $M = M_0 + \max\limits_{i\in\mathbb{S}} m(i)$, $\epsilon = \min\limits_{i\in\mathbb{S}}(\boldsymbol{P}^M)_{ij_0}$, 运用 (2.2.3) 式即得待证的结论. □

3.2 没有 Doeblin 条件的遍历理论

在这一节, 我们将看到在多大程度上, 第二章中满足 Doeblin 条件的 Markov 链的结果能被推广到不满足 Doeblin 条件的情形. 这里采用与第二章讨论相反的顺序. 也就是, 从最一般但最弱的结果开始, 然后再改进它.

3.2.1. 矩阵的收敛性: 因为我们将考虑可能有无穷多个元素的矩阵的幂级数, 所以有必要明确一下所讨论的矩阵类是什么以及在什么意义

[1] 下面结论中的 "当" 部分依赖于无穷多个素数的存在性.

下矩阵级数收敛. 为了我们的需要, 最自然的矩阵类是由下述范数有限的 $\boldsymbol{M}$ 组成的:

$$\|\boldsymbol{M}\|_{u,v} \equiv \sup_{i\in\mathbb{S}} \sum_{j\in\mathbb{S}} |(\boldsymbol{M})_{ij}|. \tag{3.2.1}$$

所有这种矩阵组成的集合记为 $M_{u,v}(\mathbb{S})$. 简单的计算表明 $M_{u,v}(\mathbb{S})$ 是 $\mathbb{R}$ 上的向量空间, $\|\cdot\|_{u,v}$ 是 $M_{u,v}(\mathbb{S})$ 上一个好的范数, 即

$$\|\boldsymbol{M}\|_{u,v} = 0 \text{ 当且仅当 } \boldsymbol{M} = 0,$$
$$\|\alpha\boldsymbol{M}\|_{u,v} = |\alpha|\|\boldsymbol{M}\|_{u,v}, \quad \alpha \in \mathbb{R}$$

和

$$\|\boldsymbol{M} + \boldsymbol{M}'\|_{u,v} \leqslant \|\boldsymbol{M}\|_{u,v} + \|\boldsymbol{M}'\|_{u,v}.$$

稍微不太明显的是

$$(\boldsymbol{M}, \boldsymbol{M}') \in M_{u,v}(\mathbb{S})^2 \implies \begin{array}{l} \boldsymbol{M}\boldsymbol{M}' \text{ 存在且} \\ \|\boldsymbol{M}\boldsymbol{M}'\|_{u,v} \leqslant \|\boldsymbol{M}\|_{u,v}\|\boldsymbol{M}'\|_{u,v}. \end{array} \tag{3.2.2}$$

为证明该结论, 首先注意到, 因为

$$\sum_k |(\boldsymbol{M})_{ik}||(\boldsymbol{M}')_{kj}| \leqslant \left(\sum_k |(\boldsymbol{M})_{ik}|\right) \sup_k |(\boldsymbol{M}')_{kj}| < \infty,$$

所以下面和式

$$(\boldsymbol{M}\boldsymbol{M}')_{ij} = \sum_k (\boldsymbol{M})_{ik}(\boldsymbol{M}')_{kj}$$

是绝对收敛的. 另外, 对于每个 i,

$$\begin{aligned} \sum_j |(\boldsymbol{M}\boldsymbol{M}')_{ij}| &\leqslant \sum_j \sum_k |(\boldsymbol{M})_{ik}||(\boldsymbol{M}')_{kj}| \\ &= \sum_k |(\boldsymbol{M})_{ik}| \left(\sum_j |(\boldsymbol{M}')_{kj}|\right) \\ &\leqslant \left(\sum_k |(\boldsymbol{M})_{ik}|\right) \|\boldsymbol{M}'\|_{u,v}, \end{aligned}$$

因此, (3.2.2) 中的不等式成立.

下面证明 $\|\cdot\|_{u,v}$ 在 $M_{u,v}(\mathbb{S})$ 上确定的度量是完备的. 为此, 需要验证: 如果 $\{\boldsymbol{M}_n\}_0^\infty \subseteq M_{u,v}(\mathbb{S})$, 则

$$\begin{aligned} &M_{ij} = \lim_{n\to\infty} (\boldsymbol{M}_n)_{ij} \text{ 对任意 } (i,j) \text{ 成立} \\ &\Longrightarrow \sup_i \sum_j |M_{ij}| \leqslant \liminf_{n\to\infty} \|\boldsymbol{M}_n\|_{u,v}. \end{aligned} \tag{3.2.3}$$

事实上, 由 Fatou 引理和定理 6.1.10, 对任意 $i \in \mathbb{S}$,

$$\sum_{j\in\mathbb{S}} |M_{ij}| \leqslant \liminf_{n\to\infty} \sum_{j\in\mathbb{S}} |(\boldsymbol{M}_n)_{ij}|,$$

因此 (3.2.3) 成立.

有了 (3.2.3), 完备性的证明可以如下进行. 假设 $\{\boldsymbol{M}_n\}_0^\infty \subseteq M_{u,v}(\mathbb{S})$ 是 $\|\cdot\|_{u,v}$-Cauchy 收敛序列: $\lim\limits_{m\to\infty} \sup\limits_{n>m} \|\boldsymbol{M}_n - \boldsymbol{M}_m\|_{u,v} = 0$. 显然, 对每对 $(i,j) \in \mathbb{S}^2$, $\{(\boldsymbol{M}_n)_{ij}\}_0^\infty$ 是 $\mathbb{R}$ 上的 Cauchy 收敛序列. 因此, 存在一个矩阵 $\boldsymbol{M}$, 使得对每一对 (i,j), $(\boldsymbol{M})_{ij} = \lim\limits_{n\to\infty} (\boldsymbol{M}_n)_{ij}$. 进一步, 由 (3.2.3) 知,

$$\|\boldsymbol{M} - \boldsymbol{M}_m\|_{u,v} \leqslant \liminf_{n\to\infty} \|\boldsymbol{M}_n - \boldsymbol{M}_m\|_{u,v} \leqslant \sup_{n>m} \|\boldsymbol{M}_n - \boldsymbol{M}_m\|_{u,v}.$$

因此, $\|\boldsymbol{M} - \boldsymbol{M}_m\|_{u,v} \to 0$.

3.2.2. Abel 收敛性: 正如在引言中所说的那样, 我们从最弱形式的收敛性结果开始. 也就是说, 我们并不试图证明 $(\boldsymbol{P}^n)_{ij}$ 的收敛性, 或者甚至当 $n \to \infty$ 时 Césaro 均值 $\frac{1}{n}\sum\limits_{m=0}^{n-1}(\boldsymbol{P}^m)_{ij}$ 的收敛性. 我们将从研究 Abel 和 $(1-s)\sum\limits_{m=0}^{\infty} s^m(\boldsymbol{P}^m)_{ij}$, $s \nearrow 1$ 的收敛性开始.

如果一个有界序列 $\{x_n\}_0^\infty \subseteq \mathbb{R}$ 满足

$$\lim_{s\nearrow 1}(1-s)\sum_{n=1}^{\infty} s^n x_n = x,$$

那么称 $\{x_n\}_0^\infty$ *Abel*收敛到 x.

很明显, Abel 收敛性比通常的收敛性更弱 (即由通常的收敛性推

出).[2] 事实上, 如果 $x_n \to x$, 那么因为 $(1-s)\sum\limits_0^\infty s^n = 1$, 故对任意 N:

$$\begin{aligned}\left|x-(1-s)\sum_{n=0}^\infty s^n x_n\right| &= (1-s)\left|\sum_{n=0}^\infty s^n(x-x_n)\right| \\ &\leqslant (1-s)\sum_{n=0}^\infty s^n|x-x_n| \\ &\leqslant N(1-s)\sup_{n\in\mathbb{N}}|x-x_n| + \sup_{n\geqslant N}|x-x_n|.\end{aligned}$$

因此, 如果 $x_n \to x$,

$$\limsup_{s\nearrow 1}\left|x-(1-s)\sum_{n=0}^\infty s^n x_n\right| \leqslant \lim_{N\to\infty}\sup_{n\geqslant N}|x-x_n| = 0.$$

另一方面, 尽管 $\{(-1)^n\}_1^\infty$ 不收敛, 但当 $s\nearrow 1$ 时,

$$(1-s)\sum_{n=0}^\infty s^n(-1)^n = (1-s)\frac{1}{1+s} \to 0.$$

也就是说, Abel 收敛性一般并不能推出通常的收敛性.

记住上面的推导, 令

$$\boldsymbol{R}(s) = (1-s)\sum_{n=0}^\infty s^n \boldsymbol{P}^n, \quad s\in[0,1). \tag{3.2.4}$$

然而, 在接受这个定义之前, 必须验证上述级数的收敛性. 为此, 首先注意到, 由于对每个 $n\geqslant 0$, $\boldsymbol{P}^n$ 是转移概率矩阵, 因此 $\|\boldsymbol{P}^n\|_{u,v}=1$. 那么, 对于 $0\leqslant m<n$,

$$\left\|(1-s)\sum_{\ell=0}^n s^\ell \boldsymbol{P}^\ell - (1-s)\sum_{\ell=0}^m s^\ell \boldsymbol{P}^\ell\right\|_{u,v} \leqslant (1-s)\sum_{\ell=m}^n s^\ell\|\boldsymbol{P}^\ell\|_{u,v} \leqslant s^m,$$

因此, 并由上面证明的完备性, (3.2.4) 中的级数关于 $\|\cdot\|_{u,v}$-范数收敛.

这里的目的是证明

$$\lim_{s\nearrow 1}(\boldsymbol{R}(s))_{ij} = \pi_{ij} \equiv \begin{cases}(\mathrm{E}[\rho_j|X_0=j])^{-1}, & 若\ i=j; \\ \mathrm{P}(\rho_j<\infty|X_0=i)\pi_{jj}, & 若\ i\neq j.\end{cases} \tag{3.2.5}$$

[2] 在习题 3.3.1 里, 将证明 Abel 收敛性也比 Césaro 收敛性更弱.

证明上式的关键是*更新方程*

$$(\boldsymbol{P}^n)_{ij} = \sum_{m=1}^{n} f(m)_{ij}(\boldsymbol{P}^{n-m})_{jj}, \quad n \geqslant 1, \tag{3.2.6}$$
其中 $f(m)_{ij} \equiv \mathrm{P}(\rho_j = m|X_0 = i)$.

这是 (2.1.13) 的一个初等应用:

$$\begin{aligned}(\boldsymbol{P}^n)_{ij} &= \sum_{m=1}^{n} \mathrm{P}(X_n = j, \rho_j = m|X_0 = i) \\ &= \sum_{m=1}^{n} \mathrm{P}(X_{n-m} = j|X_0 = j)\mathrm{P}(\rho_j = m|X_0 = i).\end{aligned}$$

下面, 对 $s \in [0,1)$, 令

$$\hat{f}(s)_{ij} \equiv \sum_{m=1}^{\infty} s^m f(m)_{ij} = \mathrm{E}[s^{\rho_j}|X_0 = i].$$

从 (3.2.6) 可得

$$\begin{aligned}(\boldsymbol{R}(s))_{ij} &= (1-s)\delta_{i,j} + (1-s)\sum_{n=1}^{\infty} s^n \left(\sum_{m=1}^{n} f(m)_{ij}(\boldsymbol{P}^{n-m})_{jj}\right) \\ &= (1-s)\delta_{i,j} + (1-s)\sum_{m=1}^{\infty} s^m f(m)_{ij} \left(\sum_{n=m}^{\infty} s^{n-m}(\boldsymbol{P}^{n-m})_{jj}\right) \\ &= (1-s)\delta_{i,j} + \hat{f}(s)_{ij}(\boldsymbol{R}(s))_{jj}.\end{aligned}$$

即对 $s \in [0,1)$,

$$(\boldsymbol{R}(s))_{ij} = (1-s)\delta_{i,j} + \hat{f}(s)_{ij}(\boldsymbol{R}(s))_{jj}. \tag{3.2.7}$$

给定 (3.2.7), 容易得到 (3.2.5). 即

$$(\boldsymbol{R}(s))_{jj} = \frac{1-s}{1-\hat{f}(s)_{jj}}.$$

这样, 如果 j 是瞬时的, 并且因此 $\hat{f}(1)_{jj} < 1$, 那么

$$\pi_{jj} = \lim_{s \nearrow 1}(\boldsymbol{R}(s))_{jj} = 0 = \frac{1}{\mathrm{E}[\rho_j|X_0 = j]}.$$

另一方面, 如果 j 是常返的, 那么, 因为

$$\frac{1-s^m}{1-s} = \sum_{\ell=0}^{m-1} s^\ell \nearrow m,$$

由单调收敛定理可得, 当 $s \nearrow 1$ 时,

$$\frac{1-\hat{f}(s)_{jj}}{1-s} = \sum_{m=1}^{\infty} \frac{1-s^m}{1-s} f(m)_{jj} \nearrow \sum_{m=1}^{\infty} m f(m)_{jj} = \mathrm{E}[\rho_j | X_0 = j].$$

同时, 当 $i \neq j$ 时,

$$(\boldsymbol{R}(s))_{ij} = \hat{f}(s)_{ij} (\boldsymbol{R}(s))_{jj} \to \mathrm{P}(\rho_j < \infty | X_0 = i) \pi_{jj}.$$

3.2.3. 平稳分布的结构: 如果一个概率向量 $\boldsymbol{\mu}$ 满足 $\boldsymbol{\mu} = \boldsymbol{\mu P}$, 那么称 $\boldsymbol{\mu}$ 是 $\boldsymbol{P}$-平稳的, 并且记为 $\boldsymbol{\mu} \in \mathrm{Stat}(\boldsymbol{P})$. 显然, 如果 $\boldsymbol{\mu}$ 是平稳的, 则对于任一 $s \in [0, 1)$, $\boldsymbol{\mu} = \boldsymbol{\mu R}(s)$. 因此, 由 (3.2.5) 和 Lebesgue 控制收敛定理,

$$(\boldsymbol{\mu})_j = \sum_i (\boldsymbol{\mu})_i (\boldsymbol{R}(s))_{ij} \to \sum_i (\boldsymbol{\mu})_i \pi_{ij}.$$

如果 j 是瞬时的, 则 $\pi_{ij} \equiv 0$. 另一方面, 如果 j 是常返的, 那么由定理 3.1.2, 要么当 $i \leftrightarrow j$ 时 $\pi_{ij} = 0$, 要么当 $i \not\leftrightarrow j$ 时 $\pi_{jj} = 0$. 因此, 无论哪种情况, 都成立

$$\boldsymbol{\mu} \in \mathrm{Stat}(\boldsymbol{P}) \implies (\boldsymbol{\mu})_j = \left(\sum_{i \leftrightarrow j} (\boldsymbol{\mu})_i \right) \pi_{jj}. \tag{3.2.8}$$

下面证明: 当 $(\boldsymbol{\pi}^C)_i \equiv \mathbf{1}_C(i) \pi_{ii}$ 时,

$$\pi_{jj} > 0 \text{ 和} C = \{i : i \leftrightarrow j\} \implies \boldsymbol{\pi}^C \in \mathrm{Stat}(\boldsymbol{P}). \tag{3.2.9}$$

为此, 首先注意到仅当 j 是常返的, $\pi_{jj} > 0$ 才成立. 因此, 所有的 $i \in C$ 是常返的, 且对于每个 $i \in C$ 和 $s \in (0, 1)$, $(\boldsymbol{R}(s))_{ik} > 0 \iff k \in C$. 另外, 对于所有 i, $(\boldsymbol{\pi}^C)_i = \lim\limits_{s \nearrow 1} (\boldsymbol{R})_{ji}$, 从而由 Fatou 引理得

$$\sum_i (\boldsymbol{\pi}^C)_i \leqslant \liminf_{s \nearrow 1} \sum_i (\boldsymbol{R}(s))_{ji} = 1.$$

类似地, 因为当 $s \nearrow 1$ 时,

$$\sum_{k \in C} (\boldsymbol{R}(s))_{jk} (\boldsymbol{P})_{ki} = \frac{(\boldsymbol{R}(s) - (1-s)\boldsymbol{I})_{ji}}{s} \to \pi_{ji} = (\boldsymbol{\pi}^C)_i,$$

故对任意 i,

$$(\boldsymbol{\pi}^C \boldsymbol{P})_i = \sum_{k \in C} \pi_{kk} (\boldsymbol{P})_{ki} \leqslant \liminf_{s \nearrow 1} \sum_{k \in C} (\boldsymbol{R}(s))_{jk} (\boldsymbol{P})_{ki} = (\boldsymbol{\pi}^C)_i.$$

但是, 如果对于某个 i 严格的不等号成立, 则由 Fubini 定理和定理 6.1.15, 我们有下面的矛盾

$$\sum_k (\boldsymbol{\pi}^C)_k = \sum_k (\boldsymbol{\pi}^C)_k \left(\sum_i (\boldsymbol{P})_{ki}\right)$$
$$= \sum_i \left(\sum_k (\boldsymbol{\pi}^C)_k (\boldsymbol{P})_{ki}\right) < \sum_i (\boldsymbol{\pi}^C)_i.$$

因此, $\boldsymbol{\pi}^C = \boldsymbol{\pi}^C \boldsymbol{P}$. 最后, 为了证明 $\boldsymbol{\pi}^C \in \mathrm{Stat}(\boldsymbol{P})$, 必须验证 $\sum_i (\boldsymbol{\pi}^C)_i = 1$. 然而, 我们已经验证了 $\boldsymbol{\pi}^C = \boldsymbol{\pi}^C \boldsymbol{P}$, 因此 $\boldsymbol{\pi}^C = \boldsymbol{\pi}^C \boldsymbol{R}(s)$. 由此, 再回顾我们已经证明的 $\sum_i (\boldsymbol{\pi}^C)_i \leqslant 1$, Lebesgue 控制收敛定理表明

$$0 < \pi_{jj} = \sum_i (\boldsymbol{\pi}^C)_i (\boldsymbol{R}(s))_{ij} \to \left(\sum_i (\boldsymbol{\pi}^C)_i\right) \pi_{jj},$$

这只有当 $\sum_i (\boldsymbol{\pi}^C)_i = 1$ 时才有可能.

在将上面的推导概括为定理之前, 我们来回顾以下定义. 设 A 是某线性空间中的一个子集, 如果对于所有 $a, a' \in A$ 和 $\theta \in [0,1]$, 有 $(1-\theta)a + \theta a' \in A$, 那么称 A 是一个*凸集*. 如果 $b \in A$, 并且对于某个 $\theta \in (0,1)$ 和 $a, a' \in A$, $b = (1-\theta)a + \theta a'$ 可推出 $b = a = a'$, 那么称 b 是凸集 A 的一个*极点*. 另外, 我们需要正常返的概念. 如果 $j \in \mathbb{S}$ 满足 $\mathrm{E}[\rho_j | X_0 = j] < \infty$, 那么称 j 是*正常返的*. 显然, 只有常返状态才可能是正常返的. 另一方面, (1.1.13) 式和 (1.1.15) 式结合起来说明可能存在*零常返状态*, 即那些常返但非正常返的状态.

3.2.10 定理. *$\mathrm{Stat}(\boldsymbol{P})$ 是 $\mathbb{R}^{\mathbb{S}}$ 中的一个凸子集. 此外, $\mathrm{Stat}(\boldsymbol{P}) \neq \varnothing$ 当且仅当至少存在一个正常返状态 $j \in \mathbb{S}$. 事实上, 对于任意 $\boldsymbol{\mu} \in \mathrm{Stat}(\boldsymbol{P})$, (3.2.8) 成立, 并且 $\boldsymbol{\mu}$ 是 $\mathrm{Stat}(\boldsymbol{P})$ 中的一个极点当且仅当存在一个正常返状态的互通集类 C 使得 $\boldsymbol{\mu} = \boldsymbol{\pi}^C$(参看 (3.2.9)). 特别地, 对任何瞬时状态 j, $(\boldsymbol{\mu})_j = 0$, 并且对于任何常返状态 j 和所有与 j 互通的状态 i, $(\boldsymbol{\mu})_i$ 同时为严格正的或者同时为零.*

证明: 唯一没有被证明的是, $\mathrm{Stat}(\boldsymbol{P})$ 的极点刻画以及当 j 是常返状态时最后一个结论.

由 (3.2.8), 当 j 是常返状态时最后一个结论的证明归结为, 如果 j 是正常返的并且 $i \leftrightarrow j$, 那么 i 是正常返的. 为此, 假设 j 是正常返的, 令 $C = \{i : i \leftrightarrow j\}$, 并且给定 $i \in C$. 那么对所有 $n \geqslant 0$,

有 $\boldsymbol{\pi}^C\boldsymbol{P}^n = \boldsymbol{\pi}^C$, 并由此, 通过选择 n 使得 $(\boldsymbol{P}^n)_{ji} > 0$, 我们得到 $(\boldsymbol{\pi}^C)_i \geqslant (\boldsymbol{\pi}^C)_j(\boldsymbol{P}^n)_{ji} > 0$.

为了刻画极点, 首先假设对于任意正常返状态的互通类 C, 有 $\boldsymbol{\mu} \neq \boldsymbol{\pi}^C$. 那么, 由 (3.2.8), 一定存在非互通的正常返状态 j 和 j' 使得 $(\boldsymbol{\mu})_j > 0 < (\boldsymbol{\mu})_{j'}$. 但是, 再次由 (3.2.8) 可知, $\boldsymbol{\mu} = \theta\boldsymbol{\pi}^C + (1-\theta)\boldsymbol{\nu}$, 其中 $C = \{i : i \leftrightarrow j\}$, $\theta = \sum\limits_{i\in C}(\boldsymbol{\mu})_i \in (0,1)$, $(\boldsymbol{\nu})_i$ 等于 0 或 $(1-\theta)^{-1}(\boldsymbol{\mu})_i$ 取决于 i 属于或不属于 C. 很明显, $\boldsymbol{\nu} \in \mathrm{Stat}(\boldsymbol{P})$, 并且由于 $\boldsymbol{\nu}_{j'} > 0 = (\boldsymbol{\pi}^C)_{j'}$, 所以 $\boldsymbol{\nu} \neq \boldsymbol{\pi}^C$. 因此, $\boldsymbol{\mu}$ 不可能是极点. 等价地, 每一个极点 $\boldsymbol{\mu}$ 都可写成 $\boldsymbol{\pi}^C$, 其中 C 是某正常返状态的互通类.

反过来, 给定一个正常返状态的互通类 C, 假设对于某个 $\theta \in (0,1)$ 和 $(\boldsymbol{\mu},\boldsymbol{\nu}) \in \mathrm{Stat}(\boldsymbol{P})^2$, $\boldsymbol{\pi}^C = (1-\theta)\boldsymbol{\mu} + \theta\boldsymbol{\nu}$. 那么对于所有 $i \notin C$, $(\boldsymbol{\mu})_i = 0$. 因此并由 (3.2.8), 得到 $\boldsymbol{\mu} = \boldsymbol{\pi}^C$, 作为推论又有 $\boldsymbol{\nu} = \boldsymbol{\pi}^C$. □

定理 3.2.10 的最后部分说明了*正常返性是一种互通类性质*. 事实上, 如果 j 是正常返的并且 $C = \{i : i \leftrightarrow j\}$, 则 $\boldsymbol{\pi}^C \in \mathrm{Stat}(\boldsymbol{P})$, 并且由于 $(\boldsymbol{\pi}^C)_j > 0$, 得 $\pi_{ii} = (\boldsymbol{\pi}^C)_i > 0$. 特别地, 如果一个 Markov 链是不可约的, 我们有理由说, 如果任何一个状态是正常返的, 该 Markov 链是正常返的. 保证正常返性的法则可参看习题 3.3.3.

3.2.4. 一个小的改进: 下面用 Césaro 收敛性代替 Abel 收敛性. 也即证明 (3.2.5) 能被

$$\lim_{n\to\infty}(\boldsymbol{A}_n)_{ij} = \pi_{ij}, \quad \text{其中 } \boldsymbol{A}_n = \frac{1}{n}\sum_{m=0}^{n-1}\boldsymbol{P}^m \tag{3.2.11}$$

所代替. 正如在习题 3.3.1 中所证明的那样, Césaro 收敛性一般不能由 Abel 收敛性推出. 事实上, 断言 Abel 收敛性隐含 Césaro 收敛性的一般结果是非常精致的. 因为原始的结果是由一个名叫 Tauber 的人所证明的, 所以通常称为 Tauber 定理. 很幸运的是, 这里所需要的 Tauber 定理非常直接.

下面的简单估计在我们的证明中起着关键作用:

$$\begin{gathered}\{a_m\}_0^\infty \subseteq [0,1], \quad A_n = \frac{1}{n}\sum_0^{n-1}a_\ell \\ \Longrightarrow |A_n - A_{n-m}| \leqslant \frac{m}{n}, \quad 0 \leqslant m < n.\end{gathered} \tag{3.2.12}$$

证明如下:

$$A_n - A_{n-m} = \frac{1}{n}\sum_{\ell=n-m}^{n-1} a_\ell - \frac{m}{n(n-m)}\sum_{\ell=0}^{n-m-1} a_\ell \begin{cases} \geqslant -\dfrac{m}{n}, \\ \leqslant \dfrac{m}{n}. \end{cases}$$

3.2.13 引理. 对于所有 (i,j), 有 $\varlimsup_{n\to\infty}(\boldsymbol{A}_n)_{ij} \leqslant e\pi_{jj}$. 此外, 对于任意 j 和任意子列 $\{n_\ell : \ell \geqslant 0\} \subseteq \mathbb{N}$, 有

$$\lim_{\ell\to\infty}(\boldsymbol{A}_{n_\ell})_{jj} = \alpha \implies \lim_{\ell\to\infty}(\boldsymbol{A}_{n_\ell})_{ij} = \mathrm{P}(\rho_j < \infty | X_0 = i)\alpha$$

对所有 i 成立.

证明: 为了证明第一部分, 注意到

$$\begin{aligned}(\boldsymbol{A}_n)_{ij} &\leqslant \frac{1}{n}\left(1-\frac{1}{n}\right)^{-n}\sum_{m=0}^{n-1}\left(1-\frac{1}{n}\right)^m (\boldsymbol{P}^m)_{ij} \\ &\leqslant \left(1-\frac{1}{n}\right)^{-n}\left(\boldsymbol{R}\left(1-\frac{1}{n}\right)\right)_{ij}.\end{aligned}$$

将上式与 (1.2.10) 和 (3.2.5) 相结合, 可得 $\varlimsup_{n\to\infty}(\boldsymbol{A}_n)_{ij} \leqslant e\pi_{ij} \leqslant e\pi_{jj}$.

为了处理第二部分, 利用 (3.2.6) 得到

$$(\boldsymbol{A}_n)_{ij} = \sum_{m=1}^{n-1} f(m)_{ij}\left(1-\frac{m}{n}\right)(\boldsymbol{A}_{n-m})_{jj}, \quad i \neq j.$$

因此,

$$\begin{aligned}|(\boldsymbol{A}_n)_{ij} - \mathrm{P}(\rho_j < n | X_0 = i)\alpha| &\leqslant \sum_{m=1}^{n-1} f(m)_{ij}\left(\frac{m}{n} + |(\boldsymbol{A}_{n-m})_{jj} - \alpha|\right) \\ &\leqslant 2\sum_{m=1}^{n-1}\frac{m}{n}f(m)_{ij} + |(\boldsymbol{A}_n)_{jj} - \alpha|,\end{aligned}$$

其中, 在第二个不等式中, 利用了 (3.2.12) 和 $\sum\limits_{m=1}^{\infty} f(m)_{ij} \leqslant 1$. 最后, 由 Lebesgue 控制收敛定理, 当 $n\to\infty$ 时, $\sum\limits_{0}^{n-1}\frac{m}{n}f(m)_{ij}$ 趋于 0, 由此, 取 $n = n_\ell$ 并令 $\ell\to\infty$, 可得想要的结论. □

我们现在可以完成 (3.2.11) 的证明了. 如果 $\pi_{jj} = 0$, 那么引理 3.2.13 的第一部分保证了对所有 i, $\lim\limits_{n\to\infty}(\boldsymbol{A}_n)_{ij} = 0 = \pi_{ij}$. 因此, 假设

$\pi_{jj} > 0$. 在这种情况下, 定理 3.2.10 表明 j 一定是正常返的, 并且当 $C \equiv \{i : i \leftrightarrow j\}$ 时, $\boldsymbol{\pi}^C \in \mathrm{Stat}(\boldsymbol{P})$. 特别地, $\pi_{jj} = \sum_{i \in C} (\boldsymbol{\pi}^C)_i (\boldsymbol{A}_n)_{ij}$. 同时, 如果 $\alpha^+ = \limsup_{n\to\infty} (\boldsymbol{A}_n)_{jj}$ 并且选择子序列 $\{n_\ell : \ell \geqslant 0\}$ 满足 $(\boldsymbol{A}_{n_\ell})_{jj} \to \alpha^+$, 则由引理 3.2.13 的第二部分和推论 3.1.4,

$$i \in C \implies \lim_{\ell\to\infty} (\boldsymbol{A}_{n_\ell})_{ij} = \alpha^+.$$

因此, 把这两个注记结合在一起, 可得到

$$\pi_{jj} = \lim_{\ell\to\infty} \sum_{i\in C} (\boldsymbol{\pi}^C)_i (\boldsymbol{A}_{n_\ell})_{ij} = \alpha^+ \sum_{\alpha\in C} (\boldsymbol{\pi}^C)_i = \alpha^+.$$

类似地, 如果 $\alpha^- = \liminf_{n\to\infty} (\boldsymbol{A}_n)_{jj}$, 那么 $\alpha^- = \pi_{jj}$. 因此 $\pi_{jj} > 0 \implies \lim_{n\to\infty} (\boldsymbol{A}_n)_{jj} = \pi_{jj}$, 再一次应用引理 3.2.13 的第二部分, 可证明 (3.2.11).

3.2.5. 平均遍历定理 (续): 正如我们在 §2.3.1 中能够运用定理 2.2.5 来证明定理 2.3.4 那样, 这里能利用 (3.2.11) 来证明下列平均遍历定理.

3.2.14 定理. *令 C 是一个正常返状态的互通类. 如果 $\mathrm{P}(X_0 \in C) = 1$, 则*

$$\lim_{n\to\infty} \mathrm{E}\left[\left(\frac{1}{n}\sum_{m=0}^{n-1} \mathbf{1}_{\{j\}}(X_m) - (\boldsymbol{\pi}^C)_j\right)^2\right] = 0.$$

更精确的陈述可参看习题 3.3.9 和 3.3.11.

证明: 因为 $\mathrm{P}(X_m \in C$ 对所有 $m \in \mathbb{N})=1$, 不失一般性, 我们可以并将假设 C 就是整个状态空间. 记住这一点, 令 $\boldsymbol{\pi} = \boldsymbol{\pi}^C$.

注意到, 如果 $\mu_i = \mathrm{P}(X_0 = i)$, 则

$$\begin{aligned}&\mathrm{E}\left[\left(\frac{1}{n}\sum_{m=0}^{n-1} \mathbf{1}_{\{j\}}(X_m) - \pi_{jj}\right)^2\right] \\ &= \sum_{i\in\mathbb{S}} \mu_i \mathrm{E}\left[\left(\frac{1}{n}\sum_{m=0}^{n-1} \mathbf{1}_{\{j\}}(X_m) - \pi_{jj}\right)^2 \middle| X_0 = i\right],\end{aligned}$$

因此, 只要证明对每个 $i \in \mathbb{S}$,

$$\lim_{n\to\infty} \mathrm{E}\left[\left(\frac{1}{n}\sum_{m=0}^{n-1} \mathbf{1}_{\{j\}}(X_m) - \pi_{jj}\right)^2 \middle| X_0 = i\right] = 0.$$

但是, 由于对所有 $i \in \mathbb{S}$, $\pi_{ii} > 0$ 以及

$$\begin{aligned} &\mathrm{E}\left[\left(\frac{1}{n}\sum_{m=0}^{n-1}\mathbf{1}_{\{j\}}(X_m) - \pi_{jj}\right)^2 \middle| X_0 = i\right] \\ &\leqslant \frac{1}{\pi_{ii}}\sum_{k\in\mathbb{S}}\pi_{kk}\mathrm{E}\left[\left(\frac{1}{n}\sum_{m=0}^{n-1}\mathbf{1}_{\{j\}}(X_m) - \pi_{jj}\right)^2 \middle| X_0 = k\right], \end{aligned}$$

所以只要证明当 $\boldsymbol{\pi}$ 是 Markov 链的初始分布时结论成立即可. 因此, 从现在开始, 作这一假设并令 $C = \mathbb{S}$.

现在令 $\boldsymbol{f}$ 是列向量, 它的第 i 个元素是 $\mathbf{1}_{\{j\}}(i) - \pi_{jj}$. 那么, 如同定理 2.3.4 的证明一样,

$$\mathrm{E}\left[\left(\frac{1}{n}\sum_{m=0}^{n-1}\mathbf{1}_{\{j\}}(X_m) - \pi_{jj}\right)^2\right] \leqslant \frac{2}{n^2}\sum_{k=0}^{n-1}(n-k)\mathrm{E}[(\boldsymbol{f})_{X_k}(\boldsymbol{A}_{n-k}\boldsymbol{f})_{X_k}].$$

因为 $\boldsymbol{\pi} \in \mathrm{Stat}(\boldsymbol{P})$, 所以

$$\mathrm{E}[(\boldsymbol{f})_{X_k}(\boldsymbol{A}_{n-k}\boldsymbol{f})_{X_k}] = \boldsymbol{\pi}(f\boldsymbol{A}_{n-k}\boldsymbol{f}),$$

并且因此前式变成了

$$\mathrm{E}\left[\left(\frac{1}{n}\sum_{m=0}^{n-1}\mathbf{1}_{\{j\}}(X_m) - \pi_{jj}\right)^2\right] \leqslant \frac{2}{n^2}\sum_{m=1}^{n} m\boldsymbol{\pi}(f\boldsymbol{A}_m\boldsymbol{f}),$$

其中 $(f\boldsymbol{A}_m\boldsymbol{f})_i \equiv (\boldsymbol{f})_i(\boldsymbol{A}_m\boldsymbol{f})_i$.

最后, 由 (3.2.11) 知, 对任何 $\varepsilon > 0$, 存在 $N_\varepsilon \in \mathbb{Z}^+$ 使得对所有 $n \geqslant N_\varepsilon$,

$$|\boldsymbol{\pi}(f\boldsymbol{A}_n\boldsymbol{f})| \leqslant \sum_i (\boldsymbol{\pi})_i|(\boldsymbol{A}_n)_{ij} - \pi_{jj}| < \varepsilon.$$

因此,

$$\begin{aligned} &\limsup_{n\to\infty}\mathrm{E}\left[\left(\frac{1}{n}\sum_{m=0}^{n-1}\mathbf{1}_{\{j\}}(X_m) - \pi_{jj}\right)^2\right] \\ &\leqslant \limsup_{n\to\infty}\frac{2}{n^2}\sum_{m=1}^{N_\varepsilon} m|\boldsymbol{\pi}(f\boldsymbol{A}_m\boldsymbol{f})| + \limsup_{n\to\infty}\frac{2\varepsilon}{n^2}\sum_{m=N_\varepsilon+1}^{n} m \leqslant \varepsilon. \quad \square \end{aligned}$$

3.2.6. 非周期情形的一个改进: 这一节的目的是证明 (参看 (3.2.5)):

$$\text{如果 } j \text{ 是瞬时的或非周期的, 则对所有 } i \in \mathbb{S},\ \lim_{n\to\infty} (\boldsymbol{P}^n)_{ij} = \pi_{ij}. \tag{3.2.15}$$

当然, 当 j 是瞬时状态时, 证明是不费力气的. 由 (2.3.7) 知,

$$j \text{ 是瞬时的} \implies \sum_{n=0}^{\infty} (\boldsymbol{P}^n)_{ij} \leqslant \mathrm{E}[T_j | X_0 = j] < \infty,$$

因此 $\lim_{n\to\infty} (\boldsymbol{P}^n)_{ij} = 0$. 同时, 因为当 j 是瞬时状态时, $\mathrm{P}(\rho_j = \infty | X_0 = j) > 0$, 所以 $\pi_{ij} \leqslant \pi_{jj} = 0$. 这样, 从现在开始, 将集中讨论 j 是常返且非周期的情形.

首先注意到 (参看 §2.3.2):

$$\begin{aligned}&\text{如果 } j \text{ 是非周期的, 则存在 } N \in \mathbb{Z}^+ \text{ 使得} \\ &\text{对所有 } n \geqslant N, \quad \max_{1 \leqslant m \leqslant n} \mathrm{P}(\rho_j^{(m)} = n | X_0 = j) > 0.\end{aligned} \tag{3.2.16}$$

为了验证这一点, 利用 (3.1.14) 给出一个 $N \in \mathbb{Z}^+$ 使得对所有 $n \geqslant N$, $(\boldsymbol{P}^n)_{jj} > 0$. 于是, 因为对于 $n \geqslant 1$,

$$(\boldsymbol{P}^n)_{jj} = \sum_{m=1}^{n} \mathrm{P}(\rho_j^{(m)} = n | X_0 = j),$$

所以 (3.2.16) 显然成立.

第二步, 也是关键的一步, 蕴涵在下列引理中.

3.2.17 引理. *假设 j 是非周期且常返的, 令 $\alpha_j^- = \liminf_{n\to\infty} (\boldsymbol{P}^n)_{jj}$ 和 $\alpha_j^+ = \limsup_{n\to\infty} (\boldsymbol{P}^n)_{jj}$. 则存在子列 $\{n_\ell^- : \ell \geqslant 1\}$ 和 $\{n_\ell^+ : \ell \geqslant 1\}$ 使得对所有 $r \geqslant 0$,*

$$\alpha_j^{\pm} = \lim_{\ell\to\infty} (\boldsymbol{P}^{n_\ell^{\pm} - r})_{jj}.$$

证明: 选择一个子列 $\{n_\ell : \ell \geqslant 1\}$ 使得 $(\boldsymbol{P}^{n_\ell})_{jj} \to \alpha_j^+$, 并应用 (3.2.16), 选择 $N \geqslant 1$ 使得对所有 $n \geqslant N$, $\max_{1 \leqslant m \leqslant n} \mathrm{P}(\rho_j^{(m)} = n | X_0 = j) > 0$. 给定 $r \geqslant N$, 选择 $1 \leqslant m \leqslant r$ 使得 $\delta \equiv \mathrm{P}(\rho_j^{(m)} = r | X_0 = j) > 0$. 现在, 对于任意 $M \in \mathbb{Z}^+$, 注意到当 $n_\ell \geqslant M + r$ 时, $(\boldsymbol{P}^{n_\ell})_{jj}$ 等于

$$\begin{aligned}&\mathrm{P}(X_{n_\ell} = j, \rho_j^{(m)} = r | X_0 = j) + \mathrm{P}(X_{n_\ell} = j, \rho_j^{(m)} \neq r | X_0 = j) \\ &= \delta (\boldsymbol{P}^{n_\ell - r})_{jj} + \mathrm{P}(X_{n_\ell} = j, n_\ell - M \geqslant \rho_j^{(m)} \neq r | X_0 = j) + \\ &\quad \mathrm{P}(X_{n_\ell} = j, \rho_j^{(m)} > n_\ell - M | X_0 = j).\end{aligned}$$

进一步有,

$$\begin{aligned}&\mathrm{P}(X_{n_\ell}=j,n_\ell-M\geqslant\rho_j^{(m)}\neq r|X_0=j)\\&=\sum_{k=1,k\neq r}^{n_\ell-M}\mathrm{P}(\rho_j^{(m)}=k|X_0=j)(\boldsymbol{P}^{n_\ell-k})_{jj}\leqslant(1-\delta)\sup_{n\geqslant M}(\boldsymbol{P}^n)_{jj},\end{aligned}$$

而

$$\mathrm{P}(X_{n_\ell}=j,\rho_j^{(m)}>n_\ell-M|X_0=j)\leqslant\mathrm{P}(\rho_j^{(m)}>n_\ell-M|X_0=j).$$

因此, 并因 j 是常返的, 故有 $\mathrm{P}(\rho_j^{(m)}<\infty|X_0=j)=1$, 在令 $\ell\to\infty$ 之后, 得

$$\alpha_j^+\leqslant\delta\liminf_{\ell\to\infty}(\boldsymbol{P}^{n_\ell-r})_{jj}+(1-\delta)\sup_{n\geqslant M}(\boldsymbol{P}^n)_{jj}.$$

因为上式对所有的 $M\geqslant 1$ 都成立, 得证

$$\alpha_j^+\leqslant\delta\liminf_{\ell\to\infty}(\boldsymbol{P}^{n_\ell-r})_{jj}+(1-\delta)\alpha_j^+,$$

由此推出 $\liminf_{\ell\to\infty}(\boldsymbol{P}^{n_\ell-r})_{jj}\geqslant\alpha_{jj}^+$. 但是, 显然有 $\limsup_{\ell\to\infty}(\boldsymbol{P}^{n_\ell-r})_{jj}\leqslant\alpha_j^+$, 因此, 我们已经证明了对所有 $r\geqslant N$, $\lim_{\ell\to\infty}(\boldsymbol{P}^{n_\ell-r})_{jj}=\alpha_j^+$. 现在选择 L 使得 $n_L\geqslant N$, 取 $n_\ell^+=n_{\ell+L}-N$, 可得对所有 $r\geqslant 0$, $\lim_{\ell\to\infty}(\boldsymbol{P}^{n_\ell^+-r})_{jj}=\alpha_j^+$.

$\{n_\ell^-:\ell\geqslant 1\}$ 的构造本质上是相同的, 留做习题. □

3.2.18 引理. 如果 j 是非周期且常返的, 则 $\limsup_{n\to\infty}(\boldsymbol{P}^n)_{jj}\leqslant\pi_{jj}$. 进而, 如果子列 $\{n_\ell^\pm:\ell\geqslant 1\}$ 是引理 3.2.17 中所给出的, 则对所有满足 $i\leftrightarrow j$ 的 i, 有 $\lim_{\ell\to\infty}(\boldsymbol{P}^{n_\ell^\pm})_{ij}=\alpha_{jj}^\pm$.

证明: 为了证明第二个论断, 只要注意到, 由引理 3.2.17 和 Lebesgue 控制收敛定理,

$$(\boldsymbol{P}^{n_\ell^\pm})_{ij}=\sum_{r=1}^{n_\ell^\pm}\mathrm{P}(\rho_j=r|X_0=i)(\boldsymbol{P}^{n_\ell^\pm-r})_{jj}\to\mathrm{P}(\rho_j<\infty|X_0=i)\alpha_j^\pm.$$

转到第一个结论, 再一次使用引理 3.2.17 的结果得到, 对于所有 $N\geqslant 1$,

$$\alpha_j^+\sum_{r=1}^N\mathrm{P}(\rho_j\geqslant r|X_0=j)=\lim_{\ell\to\infty}\sum_{r=1}^N\mathrm{P}(\rho_j\geqslant r|X_0=j)(\boldsymbol{P}^{n_\ell^+-r})_{jj}.$$

因此, 如果证明了

$$\sum_{r=1}^{N} \mathrm{P}(\rho_j \geqslant r | X_0 = j)(\boldsymbol{P}^{n-r})_{jj} \leqslant 1, \quad n \geqslant N \geqslant 1, \tag{$*$}$$

则有

$$\alpha_j^+ \mathrm{E}[\rho_j | X_0 = j] = \alpha_j^+ \sum_{r=1}^{\infty} \mathrm{P}(\rho_j \geqslant r | X_0 = j) \leqslant 1,$$

这等价于第一个结论.

为了证明 $(*)$, 注意到对任何 $n \geqslant 1$,

$$\begin{aligned}
&(\boldsymbol{P}^n)_{jj} \\
&= \sum_{r=1}^{n} \mathrm{P}(\rho_j = r | X_0 = j)(\boldsymbol{P}^{n-r})_{jj} \\
&= \sum_{r=1}^{n} \mathrm{P}(\rho_j \geqslant r | X_0 = j)(\boldsymbol{P}^{n-r})_{jj} - \sum_{r=1}^{n} \mathrm{P}(\rho_j \geqslant r+1 | X_0 = j)(\boldsymbol{P}^{n-r})_{jj} \\
&= \sum_{r=1}^{n} \mathrm{P}(\rho_j \geqslant r | X_0 = j)(\boldsymbol{P}^{n-r})_{jj} - \sum_{r=2}^{n+1} \mathrm{P}(\rho_j \geqslant r | X_0 = j)(\boldsymbol{P}^{n+1-r})_{jj},
\end{aligned}$$

由此, 并因 $\mathrm{P}(\rho_j \geqslant 1 | X_0 = j) = 1$, 对所有 $n \geqslant 1$,

$$\sum_{r=1}^{n+1} \mathrm{P}(\rho_j \geqslant r | X_0 = j)(\boldsymbol{P}^{n+1-r})_{jj} = \sum_{r=1}^{n} \mathrm{P}(\rho_j \geqslant r | X_0 = j)(\boldsymbol{P}^{n-r})_{jj}.$$

但是, 当 $n = 1$ 时, $\sum_{r=1}^{n} \mathrm{P}(\rho_j \geqslant r | X_0 = j)(\boldsymbol{P}^{n-r})_{jj} = 1$, 因此我们证明了对于所有的 $n \geqslant N \geqslant 1$,

$$\begin{aligned}
&\sum_{r=1}^{N} \mathrm{P}(\rho_j \geqslant r | X_0 = j)(\boldsymbol{P}^{n-r})_{jj} \\
&\leqslant \sum_{r=1}^{n} \mathrm{P}(\rho_j \geqslant r | X_0 = j)(\boldsymbol{P}^{n-r})_{jj} = 1. \quad \square
\end{aligned}$$

当 j 是常返且非周期时, 我们现在可以来完成 (3.2.15) 式的证明了. 由引理 3.2.18 的第一部分知, 如果 $\pi_{jj} = 0$, 那么 $\lim_{n\to\infty} (\boldsymbol{P}^n)_{jj} = 0$. 因此, 如果 $\pi_{jj} = 0$, 则对任意 i,

$$\lim_{n\to\infty} (\boldsymbol{P}^n)_{ij} = \lim_{n\to\infty} \sum_{r=1}^{n} \mathrm{P}(\rho_j = r | X_0 = i)(\boldsymbol{P}^{n-r})_{jj} = 0 = \pi_{ij}.$$

为了讨论 j 是正常返的情况, 令 $C = \{i : i \leftrightarrow j\}$, 并相应地选取 $\boldsymbol{\pi}^C$. 则 $\boldsymbol{\pi}^C \in \mathrm{Stat}(\boldsymbol{P})$. 特别地, 由引理 3.2.18 的最后一部分结果和 Lebesgue 控制收敛定理,

$$\pi_{jj} = \sum_{i \in C} (\boldsymbol{\pi}^C)_i (\boldsymbol{P}^{n_\ell^\pm})_{ij} \to \alpha_j^\pm \sum_{i \in C} (\boldsymbol{\pi}^C)_i = \alpha_j^\pm.$$

因此 $(\boldsymbol{P}^n)_{jj} \to \pi_{jj}$. 最后, 如果 $i \neq j$, 再次由 Lebesgue 定理得

$$(\boldsymbol{P}^n)_{ij} = \sum_{r=1}^{n} \mathrm{P}(\rho_j = r | X_0 = i)(\boldsymbol{P}^{n-r})_{jj} \to \mathrm{P}(\rho_j < \infty | X_0 = i)\pi_{jj} = \pi_{ij}.$$

3.2.7. 周期性结构: 上面的结果可以给出一个即使在周期不为 1 的情况下的更细致的分析. 考虑 $\mathbb{S}$ 上不可约的、常返的并具有转移概率矩阵 $\boldsymbol{P}$ 的 Markov 链, 假设周期为 $d \geqslant 2$. 这一节的基本结果是, 可以把 $\mathbb{S}$ 分解成子集 $\mathbb{S}_r$, $0 \leqslant r < d$, 并具有下列性质:

(1) $(\boldsymbol{P}^{md+r})_{jk} > 0 \implies r(k) - r(j) = r \bmod d$,

(2) 对所有充分大的 $m \geqslant 1$, $r(k) - r(j) = r \bmod d \implies (\boldsymbol{P}^{md+r})_{jk} > 0 < (\boldsymbol{P}^{md-r})_{kj}$,

(3) 对每个 $0 \leqslant r < d$, $\boldsymbol{P}^d$ 在 $\mathbb{S}_r$ 上的限制是一个非周期的、常返的、不可约的转移概率矩阵.

其中, $r(j)$ 表示满足 $j \in \mathbb{S}_r$ 的 $0 \leqslant r < d$.

为了证明这一分解的存在性, 注意到对 $0 \leqslant r < d$,

$$\exists m \geqslant 0, \quad (\boldsymbol{P}^{md+r})_{ij} > 0 \implies \exists n \geqslant 1, \quad (\boldsymbol{P}^{nd-r})_{ji} > 0. \qquad (*)$$

事实上, 由不可约性和 Euclid 运算法则知, 存在 $m' \geqslant 0$ 和 $0 \leqslant r' < d$ 使得 $(\boldsymbol{P}^{m'd+r'})_{ji} > 0$. 进而, 由 (3.1.13) 式, 对所有充分大的 m'',

$$(\boldsymbol{P}^{(m'+m'')d+r'})_{ji} \geqslant (\boldsymbol{P}^{m'd+r'})_{ji} (\boldsymbol{P}^{m''d})_{ii} > 0.$$

因此, 可以假设 $m' \geqslant 1$. 但这时 $(\boldsymbol{P}^{(m+m')d+(r+r')})_{ii} > 0$, 因此 $d|(r+r')$. 从而并注意到 $0 \leqslant r, r' < d$, 所以如果 $r = 0$, 就有 $r' = 0$; 如果 $r \geqslant 2$, 就有 $r' = d - r$.

从 $(*)$ 开始, 容易看出, 对于每一对 $(i,j) \in \mathbb{S}^2$, 存在唯一的 $0 \leqslant r < d$ 使得对于某个 $m \geqslant 0$, $(\boldsymbol{P}^{md+r})_{ij} > 0$. 这就是说, 假设对某 $m, m' \in \mathbb{N}$ 和 $0 \leqslant r, r' < d$, $(\boldsymbol{P}^{md+r})_{ij} > 0 < (\boldsymbol{P}^{m'd+r'})_{ij}$, 那么, 由 $(*)$ 知, 存在 $n \geqslant 1$ 使得 $(\boldsymbol{P}^{nd-r})_{ji} > 0$, 并由此得 $(\boldsymbol{P}^{(m'+n)d+(r'-r)})_{ii} > 0$. 因为这意味着 $d|(r'-r)$, 所以我们证明了 $r = r'$.

现在, 令 i_0 是 $\mathbb{S}$ 中一个固定的参考点, 对每个 $0 \leqslant r < d$, 定义 $\mathbb{S}_r$ 是下列 j 的集合: 存在 $m \geqslant 0$ 使得 $(\boldsymbol{P}^{md+r})_{i_0 j} > 0$. 由上面的推理知, $\mathbb{S}_r$ 是互不相交的. 另外, 由不可约性和 Euclid 运算法则, $\mathbb{S} = \bigcup_{r=0}^{d-1} \mathbb{S}_r$. 现在转向性质 (1) 的证明, 利用 (*) 来选择 $n \geqslant 0$ 和 $n' \geqslant 1$ 使得 $(\boldsymbol{P}^{nd+r(j)})_{i_0 j} > 0 < (\boldsymbol{P}^{n'd-r(k)})_{k i_0}$. 那么

$$(\boldsymbol{P}^{(n+m+n')d+(r(j)+r-r(k))})_{i_0 i_0} > 0,$$

由此得 $d|(r(j)+r-r(k))$. 等价地, $r(k)-r(j) = r \bmod d$. 反过来, 如果 $r(k)-r(j) = r \bmod d$, 选取 $n \geqslant 0$ 和 $n' \geqslant 1$ 使得 $(\boldsymbol{P}^{nd+r(k)})_{i_0 k} > 0$ 和 $(\boldsymbol{P}^{n'd-r(j)})_{j i_0} > 0$. 那么, 对任何满足 $(\boldsymbol{P}^{md})_{i_0 i_0} > 0$ 的 $m \geqslant 1$, 有 $(\boldsymbol{P}^{(n+m+n')d+r})_{jk} > 0$. 因为由 (3.1.13) 式知, 对所有充分大的 m, 有 $(\boldsymbol{P}^{md})_{i_0 i_0} > 0$, 这完成了 (2) 中左边不等式的证明. 同样可以证明 (2) 中右边的不等式. 最后, 为了验证 (3), 注意到由 (1) 可知, $\boldsymbol{P}^d$ 在 $\mathbb{S}_r$ 上的限制是一个概率转移矩阵, 并由 (2) 知, 它是不可约的和非周期的.

上述分解的存在性有着一些有趣的推论. 首先, 它表明 Markov 链以循环的方式历经状态空间: 如果它从 i 开始, 则 n 步以后, 它处于 $\mathbb{S}_{r(i)+n}$ 中, 其中下标应该被解释为模 d. 事实上, 运用这种加法约定, 我们有

$$\boldsymbol{P}^n \mathbf{1}_{\mathbb{S}_r} = \mathbf{1}_{\mathbb{S}_{r+n}}. \tag{3.2.19}$$

为了证明这一结论, 注意到, 一方面, 除非 $i \in \mathbb{S}_{r+n}$, $\boldsymbol{P}^n \mathbf{1}_{\mathbb{S}_r}(i) = 0$; 而另一方面, $\sum_{r'=0}^{d-1} \boldsymbol{P}^n \mathbf{1}_{\mathbb{S}_{r'}} = \mathbf{1}$. 因此 $i \notin \mathbb{S}_{r+n} \implies (\boldsymbol{P}^n \mathbf{1}_{\mathbb{S}_r})_i = 0$, $i \in \mathbb{S}_{r+n} \implies 1 = \sum_{r'=0}^{d-1} (\boldsymbol{P}^n \mathbf{1}_{\mathbb{S}_{r'}})_i = (\boldsymbol{P}^n \mathbf{1}_{\mathbb{S}_r})_i$.

其次, 因为对每个 $0 \leqslant r < d$, $\boldsymbol{P}^d$ 在 $\mathbb{S}_r$ 上的限制是一个不可约的、常返的和周期的转移概率矩阵, 所以对每个 $0 \leqslant r < d$ 和 $j \in \mathbb{S}_r$, 存在 $\pi_{jj}^{(r)} \in [0,1]$ 具有性质: 对所有 $(i,j) \in \mathbb{S}_r^2$, $(\boldsymbol{P}^{md})_{ij} \to \pi_{jj}^{(r)}$. 更一般地,

$$\lim_{m\to\infty} (\boldsymbol{P}^{md+s})_{ij} = \begin{cases} \pi_{jj}^{(r(j))}, & \text{若 } r(j)-r(i) = s \bmod d; \\ 0, & \text{其他}. \end{cases} \tag{3.2.20}$$

特别地, 如果 $(i,j) \in \mathbb{S}_r^2$, 则

$$(\boldsymbol{A}_{nd})_{ij} = \frac{1}{nd} \sum_{m=0}^{n-1} \sum_{s=0}^{d-1} (\boldsymbol{P}^{md+s})_{ij} = \frac{1}{nd} \sum_{m=0}^{n-1} (\boldsymbol{P}^{md})_{ij} \to \frac{\pi_{jj}^{(r)}}{d}.$$

因此, 因为已知 $(\boldsymbol{A}_n)_{ij} \to \pi_{jj}$, 故对 $0 \leqslant r < d$ 以及 $j \in \mathbb{S}_r$,

$$\pi_{jj}^{(r)} = d\pi_{jj}. \tag{3.2.21}$$

如果 $\boldsymbol{P}$ 在 $\mathbb{S}$ 上是正常返的, $\boldsymbol{P}^d$ 在 $\mathbb{S}_r$ 上的限制也同样是正常返的, 从而 $\sum_{j\in\mathbb{S}_r} \pi_{jj}^{(r)} = 1$. 因此, 在正常返的情况下, (3.2.21) 给出一个有趣的结论: $\boldsymbol{\pi}^{\mathbb{S}}$ 赋予每个 $\mathbb{S}_r$ 的概率为 $\frac{1}{d}$. 这些方面的其他应用可以参看习题 3.3.12 和 5.6.7.

3.3 习题

3.3.1. 正如 Césaro 收敛性比普通收敛性要严格弱 (即 Césaro 收敛性可由普通收敛性推出, 但反之不然), 在这个习题里, 我们将证明 Abel 收敛性比 Césaro 收敛性严格弱.

(a) 假设 $\{a_n\}_0^\infty \subseteq \mathbb{R}$ 的收敛半径小于或等于 1. 即 $\limsup\limits_{n\to\infty} |a_n|^{\frac{1}{n}} \leqslant 1$. 令 $R(s) = (1-s)\sum_0^\infty s^n a_n$, $s \in [0,1)$ 和 $A_n = \frac{1}{n}\sum_0^{n-1} a_m$, $n \geqslant 1$. 证明: 对于 $s \in [0,1)$, $R(s) = (1-s)^2\sum_1^\infty n s^{n-1} A_n$. 并利用它证明

$$\lim_{n\to\infty} A_n = a \in \mathbb{R} \implies \lim_{s\nearrow 1} R(s) = a.$$

(b) 对 $n \geqslant 0$, 取 $a_n = (-1)^{n+1} n$, 验证 $\{a_n\}_0^\infty$ 的收敛半径是 1, 并证明

$$\frac{1}{n}\sum_{m=0}^{n-1} a_m = \begin{cases} \frac{1}{2}, & \text{若 } n \text{ 是偶数;} \\ -\frac{1}{2} + \frac{1}{2n}, & \text{若 } n \text{ 是奇数} \end{cases}$$

$$\text{和} \quad (1-s)\sum_{m=0}^\infty s^m a_m = \frac{s(1-s)}{(1+s)^2}.$$

因此, $\{a_n\}_0^\infty$ 是 Abel 收敛于 0的, 但是 Césaro 发散的.

3.3.2. 回顾习题 1.3.12 中的排队模型. 证明 $\{Q_n : n \geqslant 0\}$ 是一个从 0 出发、$\mathbb{N}$ 值的 Markov 链, 写出该链的转移概率矩阵. 进而证明, 如果 $\mathrm{E}[B_1] > 0$, 则 0 是瞬时的; 如果 $\mathrm{E}[B_1] = 0$, 则 0 是零常返的; 如果 $\mathrm{E}[B_1] < 0$, 则 0 是正常返的. 为了处理 $\mathrm{E}[B_1] = 0$ 的情况, 可参看习题 1.3.11.

3.3.3. 这里是对正常返的一个检验. 给定一个转移概率矩阵 $\boldsymbol{P}$ 以及一个状态空间 $\mathbb{S}$ 上的元素 j, 令 $C=\{i\in\mathbb{S}:i\leftrightarrow j\}$. 假设 u 是 C 上的非负函数, 具有下面性质: 对于某个 $\varepsilon>0$,

$$u(i)\geqslant(\boldsymbol{Pu})_i+\varepsilon\quad 对所有\ i\in C\setminus\{j\}\ 成立.$$

(a) 首先证明

$$\mathrm{E}[u(X_{(n+1)\wedge\rho_j})|X_0=j]\leqslant\mathrm{E}[u(X_{n\wedge\rho_j})|X_0=j]-\varepsilon\mathrm{P}(\rho_j>n|X_0=j).$$

然后, 运用这一性质来证明 j 是正常返的.

(b) 假设 $\mathbb{S}=\mathbb{Z}$ 并且对所有 $i\in\mathbb{Z}\setminus\{0\}$, $|i|\geqslant\sum\limits_j|j|(\boldsymbol{P})_{ij}+\varepsilon$. 证明, 对于由 $\boldsymbol{P}$ 确定的 Markov 链, 0 是正常返的.

3.3.4. 考虑 $\mathbb{Z}$ 上的最近邻随机游动, 以概率 $p\in\left(\dfrac{1}{2},1\right)$ 向前移, 以概率 $q=1-p$ 向后移. 换句话说, 考虑 $\mathbb{Z}$ 上的马尔可夫链, 转移概率定义如下: 如果 $j=i+1$, $(\boldsymbol{P})_{ij}=p$; 如果 $j=i-1$, $(\boldsymbol{P})_{ij}=q$; 如果 $|j-i|\neq1$, $(\boldsymbol{P})_{ij}=0$. 显然, 该 Markov 链是不可约的, 且 §1.2.2 和 §1.2.1 的结果表明 0 是瞬时的. 因此, 可以应用习题 2.4.10 的讨论.

(a) 令 $\hat{\boldsymbol{P}}$ 是按照习题 2.4.10 中的方法当 $j_0=0$ 时从 $\boldsymbol{P}$ 中构造出来的. 利用 (1.1.12) 式, 证明

$$(\hat{\boldsymbol{P}})_{ij}=\begin{cases}p, & 若\ i\leqslant0,j=i+1\ \ 或\ i\geqslant1,j=i-1;\\ q, & 若\ i\leqslant0,j=i-1\ \ 或\ i\geqslant1,j=i+1;\\ 0, & 其他.\end{cases}$$

(b) 从习题 2.4.10, 我们知道对于由 $\hat{\boldsymbol{P}}$ 确定的 Markov 链, 0 是常返的. 进一步, 由习题 3.3.3 的 (b), 我们能证明它是正常返的. 事实上, 通过把习题 2.4.10 的 (b) 和 §1.1.4 的计算结合起来, 证明

$$\mathrm{E}^{\hat{\mathrm{P}}}[\rho_0|X_0=0]=\frac{2p}{p-q},$$

其中上标 $\hat{\mathrm{P}}$ 表示该数学期望是关于由 $\hat{\boldsymbol{P}}$ 所确定的 Markov 链取的.

(c) 因为 $\boldsymbol{P}$ 是不可约的, 所以 $\hat{\boldsymbol{P}}$ 也是不可约的. 因此, 由于 0 关于由 $\hat{\boldsymbol{P}}$ 确定的 Markov 链是正常返的, 所以存在关于 $\hat{\boldsymbol{P}}$ 的唯一平稳

概率向量. 求这个向量, 并利用它来证明

$$\mathrm{E}^{\hat{\mathrm{P}}}[\rho_j|X_0=j]=\begin{cases}\dfrac{2p}{p-q}, & 若\ j\in\{0,1\};\\ \dfrac{2pq^j}{p^j(p-q)}, & 若\ j\leqslant -1;\\ \dfrac{2p^j}{q^{j-1}(p-q)}, & 若\ j\geqslant 2.\end{cases}$$

3.3.5. 在 §1.2.4 中, 相当艰难地证明了 $\mathbb{Z}^3$ 上的最近邻对称随机游动是瞬时的. 人们可能希望 §3.1.2 中的判别法则能够帮助简化证明. 然而, 即使知道哪个函数 u 应该被用于定理 3.1.5 或 3.1.7, 证明中的计算也是相当细致的. 也就是说, 证明: 如果 $\alpha>0$ 是充分大的, 并且对于 $\boldsymbol{k}\in\mathbb{Z}^3$,

$$u(\boldsymbol{k})=\left(\alpha^2+\sum_{i=1}^{3}(\boldsymbol{k})_i^2\right)^{-\frac{1}{2}},$$

那么, 当 $\boldsymbol{u}$ 是由函数 u 决定的列向量, $\boldsymbol{P}$ 是 $\mathbb{Z}^3$ 上最近邻随机游动的转移概率矩阵时, $(\boldsymbol{P}\boldsymbol{u})_{\boldsymbol{k}}\leqslant u(\boldsymbol{k})\leqslant u(\boldsymbol{0})$. 也即

$$(\boldsymbol{P})_{\boldsymbol{k}\boldsymbol{\ell}}=\begin{cases}\dfrac{1}{6}, & 若\ \displaystyle\sum_{i=1}^{3}|(\boldsymbol{k})_i-(\boldsymbol{\ell})_i|=1;\\ 0, & 其他.\end{cases}$$

下面是一些提示.

(a) 给定 $\boldsymbol{k}\in\mathbb{Z}^3$, 令

$$M=1+\alpha^2+\sum_{i=1}^{3}(\boldsymbol{k})_i^2 \quad 和 \quad x_i=\frac{(\boldsymbol{k})_i}{M},\quad 1\leqslant i\leqslant 3.$$

证明 $(\boldsymbol{P}\boldsymbol{u})_k\leqslant u(\boldsymbol{k})$ 当且仅当

$$\left(1-\frac{1}{M}\right)^{-\frac{1}{2}}\geqslant\frac{1}{3}\sum_{i=1}^{3}\frac{(1+2x_i)^{\frac{1}{2}}+(1-2x_i)^{\frac{1}{2}}}{2(1-4x_i^2)^{\frac{1}{2}}}.$$

(b) 证明 $\left(1-\dfrac{1}{M}\right)^{-\frac{1}{2}}\geqslant 1+\dfrac{1}{2M}$ 和

$$\frac{(1+\xi)^{\frac{1}{2}}+(1-\xi)^{\frac{1}{2}}}{2}\leqslant 1-\frac{\xi^2}{8},\quad |\xi|<1.$$

从而得出, 如果

$$1+\frac{1}{2M}\geqslant\frac{1}{3}\sum_{i=1}^{3}\frac{1}{(1-4x_i^2)^{\frac{1}{2}}}-\frac{|\boldsymbol{x}|^2}{6},$$

那么 $(\boldsymbol{Pu})_{\boldsymbol{k}}\leqslant u(\boldsymbol{k})$, 其中 $|\boldsymbol{x}|=\sqrt{\sum_1^3 x_i^2}$ 是 $\boldsymbol{x}=(x_1,x_2,x_3)$ 的 Euclid 长度.

(c) 证明存在一个常数 $C<\infty$, 使得只要 $\alpha\geqslant 1$ 就有

$$\frac{1}{3}\sum_{i=1}^{3}\frac{1}{(1-4x_i^2)^{\frac{1}{2}}}\leqslant 1+\frac{2|\boldsymbol{x}|^2}{3}+C|\boldsymbol{x}|^4,$$

再结合前面的结论知, 我们可以取任意满足 $\alpha^2+1\geqslant 2C$ 的 $\alpha\geqslant 1$.

类似的计算表明, 对每个 $d\geqslant 3$ 和充分大的 $\alpha>0$, 函数 $(\alpha^2+\sum_1^d(\boldsymbol{k})_i^2)^{-\frac{d-2}{2}}$ 可用来证明 $\mathbb{Z}^d$ 中的最近邻对称随机游动是瞬时的. 为什么在 $d=2$ 时不会产生矛盾? 理由是当 $d=2$ 时, 满足 $\boldsymbol{Pu}\leqslant u$ 的非常值的非负函数一定具有形式 $\log(\alpha^2+(\boldsymbol{k})_1^2+(\boldsymbol{k})_2^2)$, 因此在 0 点处不能达到最大值.

3.3.6. 正如本章开头所说的, 我们这里所讲的收敛都是对每个 $(i,j)\in\mathbb{S}^2$, 是关于 $\{(\boldsymbol{P}^n)_{ij}:n\geqslant 0\}$ 的, 或者是 $\{(\boldsymbol{A}_n)_{ij}:n\geqslant 0\}$ 的. 然而, 正如马上会看到的那样, 刚刚所得到的逐点收敛性结论在某些情况下会得到改进.

(a) 假设 j 是正常返的, 令 $C=\{i:i\leftrightarrow j\}$. 给定满足性质 $\sum_{i\notin C}(\boldsymbol{\mu})_i=0$ 的概率向量 $\boldsymbol{\mu}$, 证明对每个 $i\in C$, 有 $(\boldsymbol{\mu}\boldsymbol{A}_n)_i\to\pi_{ii}$, 并且当 j 是非周期时, 有 $(\boldsymbol{\mu}\boldsymbol{P}^n)_i\to\pi_{ii}$.

(b) 下面是有关级数收敛性的一个有趣的结果. 即对每个 $m\in\mathbb{N}$, 令 $\{a_{m,n}:n\geqslant 0\}$ 是一列收敛到 b_m (当 $n\to\infty$ 时) 的实数列. 进一步, 假设对每个 $n\in\mathbb{N}$, 序列 $\{a_{m,n}:m\geqslant 0\}$ 是绝对可加的. 最后, 假设

$$\sum_{m=0}^{\infty}|a_{m,n}|\to\sum_{m=0}^{\infty}|b_m|<\infty,\quad n\to\infty.$$

证明

$$\lim_{n\to\infty}\sum_{m=0}^{\infty}|a_{m,n}-b_m|=0.$$

提示: 利用三角不等式, 证明

$$\big||a_{m,n}| - |b_m| - |a_{m,n} - b_m|\big| \leqslant 2|b_m|,$$

并应用 Lebesgue 控制收敛定理, 证明当 $n \to \infty$,

$$\begin{aligned}&\sum_{m=0}^{\infty} |a_{m,n} - b_m| \\ &\leqslant \left|\sum_{m=0}^{\infty}(|a_{m,n}| - |b_m|)\right| + \sum_{m=0}^{\infty}\big||a_{m,n}| - |b_m| - |a_{m,n} - b_m|\big| \to 0.\end{aligned}$$

(c) 回到 (a), 利用 (b) 以及 (a) 中的结论, 证明 $||\boldsymbol{\mu}\boldsymbol{A}_n - \boldsymbol{\pi}^C||_v \to 0$, 并且当 j 是非周期时, $||\boldsymbol{\mu}\boldsymbol{P}^n - \boldsymbol{\pi}^C||_v \to 0$. 特别地, 对每个满足 $\sum\limits_{i\in C}(\boldsymbol{\mu})_i = 1$ 的概率向量 $\boldsymbol{\mu}$, 有

$$\lim_{n\to\infty} \sup\{|\boldsymbol{\mu}\boldsymbol{P}^n\boldsymbol{f} - \boldsymbol{\pi}^C\boldsymbol{f}| : ||f||_u \leqslant 1\} = 0,$$

其中 $\boldsymbol{f}$ 是由函数 f 决定的列向量. 当然, 这远远不及 Doeblin 条件下所得到的结论. Doeblin 条件可以给出一个不依赖于 $\boldsymbol{\mu}$ 的收敛速度. 在一般情形, 不存在这种一致的收敛速度.

3.3.7. 当 C 是正常返的互通类时, 下面是对 $\boldsymbol{\pi}^C$ 的一个重要说明. 令 i 是一个常返状态, 并且对 $k \in \mathbb{S}$, 令 μ_k 表示 Markov 链从状态 i 出发, 在回到 i 之前, 访问状态 k 的平均次数:

$$\mu_k = \mathrm{E}\left[\sum_{m=0}^{\rho_i - 1} \mathbf{1}_{\{k\}}(X_m)\middle| X_0 = i\right] \in [0, \infty].$$

由 $(\boldsymbol{\mu})_k = \mu_k$ 决定行向量 $\boldsymbol{\mu} \in [0,\infty]^{\mathbb{S}}$.

(a) 证明对所有 $j \in \mathbb{S}$,

$$(\boldsymbol{\mu}\boldsymbol{P})_j = \mathrm{E}\left[\sum_{m=1}^{\rho_i} \mathbf{1}_{\{j\}}(X_m)\middle| X_0 = i\right] = \mu_j.$$

这样, 无需对 i 作进一步的假设, $\boldsymbol{\mu}$ 是 $\boldsymbol{P}$-平稳的: $\boldsymbol{\mu} = \boldsymbol{\mu}\boldsymbol{P}$.

(b) 很明显, 除非 i 是正常返的, $\mu_i = 1$ 并且 $\sum\limits_j \mu_j = \infty$. 但是可以证明, 除非 $i \leftrightarrow j$, 有 $\mu_j = 0$, 并且如果 $i \leftrightarrow j$, 那么 $\mu_j \in (0, \infty)$.

提示: 证明

$$\mathrm{P}(\rho_j^{(m)} < \rho_j | X_0 = i) = \mathrm{P}(\rho_j < \rho_i | X_0 = j)^{m-1}\mathrm{P}(\rho_j < \rho_i | X_0 = i).$$

(c) 如果 i 是正常返的, 证明

$$\bar{\boldsymbol{\mu}} \equiv \frac{\boldsymbol{\mu}}{\sum\limits_k \mu_k} = \boldsymbol{\pi}^C.$$

等价地, 当 i 是正常返时, 有

$$(\boldsymbol{\pi}^C)_j = \frac{\mathrm{E}\left[\sum\limits_{m=0}^{\rho_i-1} \mathbf{1}_{\{j\}}(X_m)\middle|X_0 = i\right]}{\mathrm{E}[\rho_i|X_0 = i]}.$$

换句话说, $(\boldsymbol{\pi}^C)_j$ *是 Markov 链在回到状态 i 之前处于状态 j 的相对平均次数.*

3.3.8. 接习题 3.3.7, 但现在假设参考点 i 是零常返的. 这样, 当 $\boldsymbol{\mu} \in [0,\infty)^{\mathbb{S}}$ 是习题 3.3.7 中引入的 $\boldsymbol{P}$-平稳测度时, $\sum\limits_j (\boldsymbol{\mu})_j = \infty$. 在这个习题中, 我们将证明, 除了相差一个常数倍外, $\boldsymbol{\mu}$ 是唯一 $\boldsymbol{P}$-平稳的 $\boldsymbol{\nu} \in [0,\infty)^{\mathbb{S}}$, 满足性质: 除非 $i \to j$ (从而 $i \leftrightarrow j$), $(\boldsymbol{\nu})_j = 0$. 等价地, 给定这样的 $\boldsymbol{\nu}$, 有 $\boldsymbol{\nu} = (\boldsymbol{\nu})_i\boldsymbol{\mu}$.

(a) 假设 $\boldsymbol{\nu} \in [0,\infty)^{\mathbb{S}}$ 满足 $\boldsymbol{\nu} = \boldsymbol{\nu}\boldsymbol{P}$. 如果 $(\boldsymbol{\nu})_i = 1$, 证明对所有 $j \in \mathbb{S}$ 和 $n \geqslant 0$, 有

$$\begin{aligned}(\boldsymbol{\nu})_j = \sum_{k \neq i} (\boldsymbol{\nu})_k \mathrm{P}(X_n = j,\ \rho_i > n|X_0 = k) +\\ \mathrm{E}\left[\sum_{m=0}^{n\wedge(\rho_i-1)} \mathbf{1}_{\{j\}}(X_m)\middle|X_0 = i\right].\end{aligned}$$

提示: 对 $n \geqslant 0$ 采用归纳法. 当 $n = 0$ 时, 显然成立. 为了进行归纳的下一步, 利用 (2.1.1) 式和 Fubini 定理证明:

$$\begin{aligned}&\sum_{k\neq i} (\boldsymbol{\nu})_k \mathrm{P}(X_n = j, \rho_i > n|X_0 = k)\\ =&\sum_{k\neq i} (\boldsymbol{\nu}\boldsymbol{P})_k \mathrm{P}(X_n = j, \rho_i > n|X_0 = k)\\ =&\sum_{\ell} (\boldsymbol{\nu})_\ell \mathrm{P}(X_{n+1} = j, \rho_i > n+1|X_0 = \ell).\end{aligned}$$

(b) 假设 $\boldsymbol{\nu} = \boldsymbol{\nu}\boldsymbol{P}$, 并且除非 $i \leftrightarrow j$, $(\boldsymbol{\nu})_j = 0$. 证明 $\boldsymbol{\nu} = (\boldsymbol{\nu})_i\boldsymbol{\mu}$.

提示: 首先证明如果 $(\boldsymbol{\nu})_i = 0$, 那么 $\boldsymbol{\nu} = 0$. 因此, 只要考虑 $(\boldsymbol{\nu})_i = 1$ 的

情形. 从 (a) 出发, 应用单调收敛定理知, 对所有 $j \in \mathbb{S}$ 有 $(\boldsymbol{\nu})_j \geqslant (\boldsymbol{\mu})_j$. 最后考察 $\boldsymbol{\omega} \equiv \boldsymbol{\nu} - \boldsymbol{\mu}$, 得出 $\boldsymbol{\omega} = \mathbf{0}$.

3.3.9. 令 C 是正常返状态的互通类. 定理 3.2.14 之所以被称为平均遍历定理, 是因为它考虑的是均方意义下收敛. 当然, 均方收敛性推出依概率收敛性, 但一般而言, 它不能推出以概率 1 收敛. 然而, 正如这里所证明的, 当 $\mathrm{P}(X_0 \in C) = 1$ 并且 $j \in C$ 时,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \mathbf{1}_{\{j\}}(X_m) = \pi_{jj} \text{ 以概率 1 成立.} \tag{3.3.10}$$

注意, 在 §2.3.3 中, 我们在 Doeblin 条件下证明了逐点遍历定理 (2.3.10), 而 (3.3.10) 表明即使没有 Doeblin 条件, 同样的**逐点遍历定理**也成立. 事实上, 存在一个很一般的结论, (3.3.10) 只是一种特殊情况. 后者最早由 G. D. Birkhoff 所证明. 不过, 我们将不按照 Birkhoff 的方法, 而是如在 §2.3.3 中所做的那样, 利用强大数律. 但这里需要有关对独立同分布、可积随机变量平均值成立的完整的结论 (参看 [9] 中的定理 1.4.11).

(a) 证明: 只要对某个 $i \in C$, $\mathrm{P}(X_0 = i) = 1$ 时结论成立即可.

(b) 令 $\rho_i^{(0)} = 0$, 用 $\rho_i^{(m)}$ 表示第 m 次回到状态 i 的时刻. 如果

$$\tau_m = \rho_i^{(m)} - \rho_i^{(m-1)} \quad \text{和} \quad Y_m = \sum_{\ell=\rho_i^{(m-1)}}^{\rho_i^{(m)}-1} \mathbf{1}_{\{j\}}(X_\ell),$$

证明在 $X_0 = i$ 的条件下, $\{\tau_m, m \geqslant 1\}$ 和 $\{Y_m, m \geqslant 1\}$ 是独立同分布、非负可积的随机变量序列. 特别地, 作为强大数律和 (3.2.5) 式的应用, 再加上习题 3.3.7 中的结论得证, 在 $X_0 = i$ 的条件下,

$$\lim_{m\to\infty} \frac{\rho_i^{(m)}}{m} = \frac{1}{\pi_{ii}} \quad \text{和} \quad \lim_{m\to\infty} \frac{1}{m} \sum_{\ell=0}^{\rho_i^{(m)}-1} \mathbf{1}_{\{j\}}(X_m) = \frac{\pi_{jj}}{\pi_{ii}} \tag{$*$}$$

都以概率 1 成立. 因此, $\lim\limits_{m\to\infty} \dfrac{1}{\rho_i^{(m)}} \sum\limits_{\ell=0}^{\rho_i^{(m)}-1} \mathbf{1}_{\{j\}}(X_m) = \pi_{ij}$ 以概率 1 成立.

(c) 鉴于 (a) 和 (b) 中的结论, 只要验证在 $X_0 = i$ 的条件下,

$$\lim_{n\to\infty} \left| \frac{1}{n} \sum_{0}^{n-1} \mathbf{1}_{\{j\}}(X_\ell) - \frac{1}{\rho_i^{(m_n)}} \sum_{0}^{\rho_i^{(m_n)}-1} \mathbf{1}_{\{j\}}(X_\ell) \right| = 0$$

以概率 1 成立即可, 其中 m_n 是由 $\rho_i^{(m_n-1)} \leqslant n < \rho_i^{(m_n)}$ 确定的 $\mathbb{Z}^+$ 值随机变量. 为此, 首先证明

$$\left|\frac{1}{n}\sum_0^{n-1}\mathbf{1}_{\{j\}}(X_\ell) - \frac{1}{\rho_i^{(m_n)}}\sum_0^{\rho_i^{(m_n)}-1}\mathbf{1}_{\{j\}}(X_\ell)\right| \leqslant \frac{\tau_{m_n}}{m_n}.$$

然后, 利用 (*) 式的第一部分, 证明 $\mathrm{P}(\lim_{n\to\infty} m_n = \infty | X_0 = i) = 1$. 最后, 验证对任意 $\epsilon > 0$,

$$\begin{aligned}&\mathrm{P}\left(\sup_{m\geqslant M}\frac{\tau_m}{m} \geqslant \epsilon \middle| X_0 = i\right)\\ &\leqslant \sum_{m=M}^{\infty}\mathrm{P}(\rho_i \geqslant m\epsilon | X_0 = i) \leqslant \frac{1}{\epsilon}\mathrm{E}[\rho_i, \rho_i \geqslant M\epsilon | X_0 = i],\end{aligned}$$

并由此完成 (3.3.10) 的证明.

(d) 引入*经验测度* $\boldsymbol{L}_n$, 它是度量停留在某状态上平均时间的随机概率向量. 即 $(\boldsymbol{L}_n)_i = \frac{1}{n}\sum_0^{n-1}\mathbf{1}_{\{i\}}(X_m)$. 将这里证明的结论和习题 3.3.6 中的 (b) 结合起来得证, 当 $\mathrm{P}(X_0 \in C) = 1$ 时, $\lim_{n\to\infty}\|\boldsymbol{L}_n - \boldsymbol{\pi}^C\|_v = 0$ 以概率 1 成立.

3.3.11. 尽管习题 3.3.9 中的结论只适用于正常返状态, 但是对非正常返的状态, 存在着一个相应的极限定理. 证明: 如果 j 不是正常返的, 那么不管初始分布是什么, 都有

$$\mathrm{P}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{m=0}^{n-1}\mathbf{1}_{\{j\}}(X_m) = 0\right) = 1.$$

当 j 是瞬时状态时, $\mathrm{E}\left[\sum_0^{\infty}\mathbf{1}_{\{j\}}(X_m)\right] < \infty$, 因此结论显然成立. 当 j 是零常返时, 注意到只要考虑 $\mathrm{P}(X_0 = j) = 1$ 的情况即可. 然后, 注意到在 $X_0 = j$ 的条件下, $\{\rho_j^{(m+1)} - \rho_j^{(m)} : m \geqslant 0\}$ 是一列独立同分布、$\mathbb{Z}^+$ 值的随机变量. 应用强大数律可得, 对任意 $R \geqslant 1$, 有

$$\begin{aligned}&\mathrm{P}\left(\liminf_{m\to\infty}\frac{\rho_j^{(m)}}{m} \geqslant H(R)\middle| X_0 = j\right)\\ &\geqslant \mathrm{P}\left(\liminf_{m\to\infty}\frac{\rho_j^{(m)}\wedge R}{m} \geqslant H(R)\middle| X_0 = j\right) = 1,\end{aligned}$$

其中, 当 $R \nearrow \infty$ 时, $H(r) \equiv \frac{1}{2}\mathrm{E}[\rho_j \wedge R | X_0 = j] \nearrow \infty$. 因此, 给定

$X_0 = j$, $\dfrac{\rho_j^{(m)}}{m} \to \infty$ 以概率 1 成立. 最后, 验证对任意的 $0 < \epsilon < 1$,

$$\mathrm{P}\left(\sup_{n \geqslant N} \frac{1}{n} \sum_{0}^{n-1} \mathbf{1}_{\{j\}}(X_m) \geqslant \epsilon \,\middle|\, X_0 = j\right) \leqslant \mathrm{P}\left(\sup_{n \geqslant N} \frac{\rho_j^{([n\epsilon])}}{n} \leqslant \frac{1}{\epsilon} \,\middle|\, X_0 = j\right)$$

$$\leqslant \mathrm{P}\left(\sup_{m \geqslant N\epsilon} \frac{\rho_j^{(m)}}{m} \leqslant \frac{1}{\epsilon} \,\middle|\, X_0 = j\right),$$

结合前面的推理可得到待证的结论.

3.3.12. 当 j 关于 $\boldsymbol{P}$ 是非周期时, 对所有 $i \in \mathbb{S}$, $\lim\limits_{n\to\infty}(\boldsymbol{P}^n)_{ij}$ 存在, 并且除非 j 是正常返的, 它等于 0. 当 $d(j) > 1$ 并且 j 是正常返时, 证明 $\lim\limits_{n\to\infty}(\boldsymbol{P}^n)_{jj}$ 不存在. 另一方面, 即使 $d(j) > 1$, 证明对非正常返的 $j \in \mathbb{S}$, 有 $\lim\limits_{n\to\infty}(\boldsymbol{P}^n)_{ij} = 0$.

第四章 连续时间 Markov 过程

到目前为止, 我们处理的都是离散时间参数 $n \in \mathbb{N}$ 的 Markov 过程. 在本章中, 我们将介绍连续时间参数 $t \in [0, \infty)$ 的 Markov 过程, 但这些过程仍然在某个可数状态空间 $\mathbb{S}$ 中取值.

4.1 Poisson 过程

正如 $\mathbb{Z}^d$ 上具有独立同分布增量的 Markov 过程 (随机游动) 是最简单的离散参数的 Markov 过程一样, 最简单的连续时间 Markov 过程是那些增量独立并且齐次的过程, 即增量的分布只与产生增量的时间区间长度有关的过程. 更准确地说, 在本节中我们将处理具有下列性质的 $\mathbb{Z}^d$ 值随机过程 $\{\boldsymbol{X}(t) : t \geqslant 0\}$:

$$\mathrm{P}(\boldsymbol{X}(0) = \mathbf{0}) = 1$$

且

$$\begin{aligned} &\mathrm{P}(\boldsymbol{X}(t_1) - \boldsymbol{X}(t_0) = j_1, \cdots, \boldsymbol{X}(t_n) - \boldsymbol{X}(t_{n-1}) = j_n) \\ = &\prod_{m=1}^{n} \mathrm{P}(\boldsymbol{X}(t_m - t_{m-1}) = j_m) \end{aligned}$$

对所有 $n \geqslant 1, 0 \leqslant t_0 < \cdots < t_n$ 和 $(j_1, \cdots, j_n) \in (\mathbb{Z}^d)^n$ 成立.

4.1.1. 简单 Poisson 过程: 简单 *Poisson* 过程是这样一个 $\mathbb{N}$ 值随机过程 $\{N(t) : t \geqslant 0\}$: 从 0 (即 $N(0) = 0$) 开始, 并且在 0 处停留单位

指数时间 E_1 (即对 $0 \leqslant t < E_1$ 有 $N(t) = 0$ 和 $\mathrm{P}(E_1 > t) = e^{-t}$), 在 E_1 时刻转到 1 并且在此处有一个独立的单位指数停留时间 E_2, 同样在 $E_1 + E_2$ 时刻转到 2, 依此类推. 更准确地说, 如果 $\{E_n : n \geqslant 1\}$ 是一列相互独立, 具有单位指数的随机变量, 且有

$$J_n = \begin{cases} 0, & \text{当 } n = 0; \\ \sum_{m=1}^{n} E_m, & \text{当 } n \geqslant 1. \end{cases}$$

那么由

$$N(t) = \max\{n \geqslant 0 : J_n \leqslant t\} \tag{4.1.1}$$

定义的随机过程 $\{N(t) : t \geqslant 0\}$ 就是一个简单 Poisson 过程. 按照 (4.1.1), 由于对所有 $n \geqslant 1$, $E_n > 0$ 以及 $\sum_{m=1}^{\infty} E_m = \infty$ 以概率 1 成立, 所以轨道 $t \in [0, \infty) \mapsto N(t)$ 以概率 1 逐段为常数、右连续, 并且当它跳跃的时候总是跳 $+1$ 个单位: 即对所有 $t > 0$ 有 $N(t) - N(t-) \in \{0, 1\}$, 其中 $N(t-) \equiv \lim_{s \nearrow t} N(s)$ 是 $N(\cdot)$ 在 t 点的左极限.

下面我们将证明 $\{N(t) : t \geqslant 0\}$ 以独立齐次增量形式运动[1]. 也就是说, 我们将证明, 对每个 $s, t \in [0, \infty)$, $N(t) - N(s)$ 与 $\sigma(\{N(\tau) : \tau \in [0, s]\})$ 是独立的 (参看 §6.1.3), 并且和 $N(t)$ 具有相同的分布:

$$\mathrm{P}(N(s+t) - N(s) = n | N(\tau), \tau \in [0, s]) = \mathrm{P}(N(t) = n), \quad n \in \mathbb{N}. \tag{4.1.2}$$

等价地, 我们来验证: 当 $(s, t) \in [0, \infty)^2$, $A \in \sigma(\{N(\tau) : \tau \in [0, s]\})$ 时, 对所有 $n \in \mathbb{N}$ 有 $\mathrm{P}(\{N(s+t) - N(s) \geqslant n\} \bigcap A) = \mathrm{P}(N(t) \geqslant n)\mathrm{P}(A)$. 由于 $n = 0$ 时是显然的, 故假定 $n \in \mathbb{Z}^+$. 另外, 由于我们可以把 A 写作是不相交的 $A \bigcap \{N(s) = m\}, m \in \mathbb{N}$, 的并, 并且它们的每一个也都是在 $\sigma(\{N(\tau) : \tau \in [0, s]\})$ 里的, 所以我们假定, 在 A 上, 对某个 m, 有 $N(s) = m$. 这样, 我们可以写 $A = \{J_{m+1} > s\} \bigcap B$, 其中 $B \in \sigma(\{E_1, \cdots, E_m\})$, 并且在 B 上 $J_m \leqslant s$. 由于 $N(s+t) \geqslant m+n \iff J_{m+n} \leqslant s+t$ 以及 $\sigma(\{J_\ell - J_m : \ell > m\})$ 与 $\sigma(\{E_k : k \leqslant m\})$ 是独立的,

[1] 实际上, 这里的主要目的是给出一个推理思路, 它在后面是有用的. 简单 Poisson 过程具有独立齐次增量的一个更直接的证明放在后面的习题 4.5.1 里.

应用 (6.4.2) 式可证

$$\begin{aligned}
&\mathrm{P}(\{N(s+t)-N(s)\geqslant n\}\bigcap A)\\
&=\mathrm{P}(\{J_{m+n}\leqslant s+t\}\bigcap\{J_{m+1}>s\}\bigcap B)\\
&=\mathrm{P}(\{J_{m+n}-J_m\leqslant s+t-J_m\}\bigcap\{J_{m+1}-J_m>s-J_m\}\bigcap B)\\
&=\mathrm{E}[v(J_m),B],
\end{aligned}$$

其中, 对 $\xi\in[0,s]$ 有

$$\begin{aligned}
v(\xi)&\equiv\mathrm{P}(\{J_{m+n}-J_m\leqslant s+t-\xi\}\bigcap\{J_{m+1}-J_m>s-\xi\})\\
&=\mathrm{P}(\{J_n\leqslant s+t-\xi\}\bigcap\{E_1>s-\xi\})\\
&=\mathrm{P}(\{J_n\leqslant s+t-\xi\}\bigcap\{E_n>s-\xi\})\\
&=\mathrm{P}(\{J_{n-1}+E_n\leqslant s+t-\xi\}\bigcap\{E_n>s-\xi\})\\
&=\mathrm{E}[w(\xi,E_n),E_n>s-\xi],
\end{aligned}$$

其中 $w(\xi,\eta)\equiv\mathrm{P}(J_{n-1}\leqslant s+t-\xi-\eta)$, $\xi\in[0,s]$ 且 $\eta\in[s-\xi,s+t-\xi]$.

到目前为止, 对于指数随机变量, 除了它取正值以外, 还没有用过其他性质. 在下面部分, 我们将利用它们的特有性质, 即指数随机变量 E 具有的 “无记忆性”. 也就是说, $\mathrm{P}(E>a+b|E>a)=\mathrm{P}(E>b)$, 从而易知, 对任意非负、$\mathcal{B}_{[0,\infty)}$ 可测的函数 f, 有

$$\mathrm{E}[f(E),E>a]=e^{-a}\mathrm{E}[f(a+E)].\tag{4.1.3}$$

特别地, 我们有

$$\begin{aligned}
v(\xi)&=\mathrm{E}[w(\xi,E_n),E_n>s-\xi]=e^{-(s-\xi)}\mathrm{E}[w(\xi,E_n+s-\xi)]\\
&=e^{-(s-\xi)}\mathrm{P}(J_{n-1}\leqslant t-E_n)=e^{-(s-\xi)}\mathrm{P}(J_n\leqslant t)\\
&=e^{-(s-\xi)}\mathrm{P}(N(t)\geqslant n).
\end{aligned}$$

因而我们也就证明了

$$\mathrm{P}(\{N(s+t)-N(s)\geqslant n\}\bigcap A)=\mathrm{E}[e^{-(s-J_m)},B]\mathrm{P}(N(t)\geqslant n).$$

最后, 由于

$$\begin{aligned}
\mathrm{P}(A)&=\mathrm{P}(\{J_{m+1}>s\}\bigcap B)=\mathrm{P}(\{E_{m+1}>s-J_m\}\bigcap B)\\
&=\mathrm{E}[e^{-(s-J_m)},B],
\end{aligned}$$

(4.1.2) 式得证. 至此, 我们已经知道, *简单 Poisson 过程 $\{N(t): t \geqslant 0\}$ 具有相互独立的齐次增量*.

在继续下面的内容之前, 我们先来求 $N(t)$ 的分布. 由于 n 个相互独立、单位指数随机变量的和服从 $\Gamma(n)$ 分布, 所以

$$\begin{aligned} \mathrm{P}(N(t)=n) &= \mathrm{P}(J_n \leqslant t < J_{n+1}) = \mathrm{P}(J_n \leqslant t) - \mathrm{P}(J_{n+1} \leqslant t) \\ &= \frac{1}{(n-1)!}\int_0^t \tau^{n-1}e^{-\tau}d\tau - \frac{1}{n!}\int_0^t \tau^n e^{-\tau}d\tau = e^{-t}\frac{t^n}{n!}. \end{aligned}$$

这说明 $N(t)$ 是一个均值为 t 的 Poisson 随机变量. 结合上述结果和 (4.1.2) 式得, 对所有 $s, t \in [0,\infty)$ 和 $n \in \mathbb{N}$, 有

$$\mathrm{P}(N(s+t) - N(s) = n | N(\tau), \tau \in [0,s]) = e^{-t}\frac{t^n}{n!}. \qquad (4.1.4)$$

同样地, 再次由 (4.1.2) 式可知，$\{N(t): t \geqslant 0\}$ 具有下列 Markov 性质:

$$\mathrm{P}(N(s+t) = n | N(\tau), \tau \in [0,s]) = e^{-t}\frac{t^{n-N(s)}}{(n-N(s))!}\mathbf{1}_{[0,n]}(N(s)). \quad (4.1.5)$$

4.1.2. $\mathbb{Z}^d$ 上的复合 Poisson 过程: 有了最简单的 Poisson 过程的定义之后, 我们就能轻松地定义一类与 $\mathbb{Z}^d$ 上的随机游动类似的具有连续时间的较为复杂的过程. 也就是说, 假设 $\boldsymbol{\mu}$ 是 $\mathbb{Z}^d$ 上在原点 $\mathbf{0}$ 的质量 (概率) 为 0 的概率向量, 那么*具有跳跃分布 $\boldsymbol{\mu}$ 和速率 $R \in (0,\infty)$ 的 Poisson 过程*就是具有下述性质的随机过程 $\{\boldsymbol{X}(t): t \geqslant 0\}$: 从 $\mathbf{0}$ 点出发, 在 $\mathbf{0}$ 点处停留均值为 R^{-1} 的指数时间, 并且以概率 $(\boldsymbol{\mu})_k$ 跳跃 $k \in \mathbb{Z}^d$, 在该处停留均值为 R^{-1} 的独立的指数时间, 然后再以分布 $\boldsymbol{\mu}$ 跳跃一个随机数值, 依此类推. 因此, 简单 Poisson 过程考虑的就是当 $d=1, (\boldsymbol{\mu})_1 = 1, R = 1$ 时的情况. 特别地, 简单 Poisson 过程的跃度是确定的, 而复合 Poisson 过程的跃度一般来说是随机的.

下面我们来构造复合 Poisson 过程, 令 $\{\boldsymbol{B}_n : n \geqslant 1\}$ 是一列相互独立的具有分布 $\boldsymbol{\mu}$ 的 $\mathbb{Z}^d$ 值随机变量, 引入随机游动 $\boldsymbol{X}_0 = \mathbf{0}$, $\boldsymbol{X}_n = \sum_{m=1}^{n}\boldsymbol{B}_m, n \geqslant 1$, 并且定义 $\{\boldsymbol{X}(t): t \geqslant 0\}$ 使得 $\boldsymbol{X}(t) = \boldsymbol{X}_{N(Rt)}$, 其中 $\{N(t): t \geqslant 0\}$ 是一个独立于 $\{\boldsymbol{B}_m\}$ 的简单泊松过程. 定理 6.3.2 保证了所有这些随机变量的存在性 (参看 §1.2.1 的脚注). 显然, $\boldsymbol{X}(0) = \mathbf{0}$ 且 $t \in [0,\infty) \mapsto \boldsymbol{X}(t) \in \mathbb{Z}^d$ 是一个逐段为不变常数、右连续的 $\mathbb{Z}^d$ 值

轨道. 另外, 由于 $(\boldsymbol{\mu})_{\mathbf{0}}=0$, 显然[2], $t \rightsquigarrow \boldsymbol{X}(t)$ 在时间区间 $(s,t]$ 内引起的跳跃数正好为 $N(Rt)-N(Rs)$, 并且 $\boldsymbol{X}_n-\boldsymbol{X}_{n-1}$ 是 $t \rightsquigarrow \boldsymbol{X}(t)$ 的第 n 次跳跃高度. 因此, 如果令 $J_0\equiv 0$, 并且对 $n\geqslant 1$, J_n 表示 $t \rightsquigarrow \boldsymbol{X}(t)$ 的第 n 次跳跃的时刻, 那么 $N(Rt)=n \Longleftrightarrow J_n\leqslant Rt<J_{n+1}$ 且 $\boldsymbol{X}(J_n)-\boldsymbol{X}(J_{n-1})=\boldsymbol{X}_n-\boldsymbol{X}_{n-1}$. 也就是说, 如果 $\{E_n:n\geqslant 1\}$ 表示在 $\{N(t):t\geqslant 0\}$ 之外构建的单位指数随机变量序列, 那么 $J_n-J_{n-1}=\dfrac{E_n}{R}$, 对 $t\in(J_{n-1},J_n)$, $\boldsymbol{X}(t)-\boldsymbol{X}(t-)=\mathbf{0}$, 而且 $\boldsymbol{X}(J_n)-\boldsymbol{X}(J_{n-1})=\boldsymbol{B}_n$. 所以, $\{\boldsymbol{X}(t):t\geqslant 0\}$ 确实是一个具有跳跃分布 $\boldsymbol{\mu}$ 和速率 R 的复合 Poisson 过程.

我们下面将要证明, 复合 Poisson 过程以齐次的、相互独立的增量运动:

$$\mathrm{P}(\boldsymbol{X}(s+t)-\boldsymbol{X}(s)=\boldsymbol{k}|\boldsymbol{X}(\tau),\tau\in[0,s])=\mathrm{P}(\boldsymbol{X}(t)=\boldsymbol{k}),\quad \boldsymbol{k}\in\mathbb{Z}^d. \tag{4.1.6}$$

为了证明上述结论, 我们利用上面已经介绍过的表达式 $\boldsymbol{X}(t)=\boldsymbol{X}_{N(Rt)}$. 给定 $A\in\sigma(\{\boldsymbol{X}(\tau):\tau\in[0,s]\})$, 我们需证

$$\mathrm{P}(\{\boldsymbol{X}(s+t)-\boldsymbol{X}(s)=\boldsymbol{k}\}\cap A)=\mathrm{P}(\{\boldsymbol{X}(s+t)-\boldsymbol{X}(s)=\boldsymbol{k}\})\mathrm{P}(A).$$

回顾 (4.1.2) 式的推导过程, 不失一般性, 我们假定: 对某个 $m\in\mathbb{N}$, 在 A 上 $N(Rs)=m$. 从而, A 与 $\sigma(\{\boldsymbol{X}_{m+n}-\boldsymbol{X}_m:n\geqslant 0\}\bigcup\{N(R(s+t))-N(Rs)\})$ 独立. 因此,

$$\begin{aligned}
&\mathrm{P}(\{\boldsymbol{X}(s+t)-\boldsymbol{X}(s)=\boldsymbol{k}\}\cap A)\\
&=\sum_{n=0}^{\infty}\mathrm{P}(\{\boldsymbol{X}(s+t)-\boldsymbol{X}(s)=\boldsymbol{k},\ N(R(s+t))-N(Rs)=n\}\cap A)\\
&=\sum_{n=0}^{\infty}\mathrm{P}(\{\boldsymbol{X}_{m+n}-\boldsymbol{X}_m=\boldsymbol{k},\ N(R(s+t))-N(Rs)=n\}\cap A)\\
&=\sum_{n=0}^{\infty}\mathrm{P}(\boldsymbol{X}_n=\boldsymbol{k})\mathrm{P}(N(Rt)=n)\mathrm{P}(A)\\
&=\sum_{n=0}^{\infty}\mathrm{P}(\boldsymbol{X}_n=\boldsymbol{k},\ N(Rt)=n)\mathrm{P}(A)=\mathrm{P}(\boldsymbol{X}(t)=\boldsymbol{k})\mathrm{P}(A).
\end{aligned}$$

---

[2] 这就是我们假定 $(\boldsymbol{\mu})_{\mathbf{0}}=0$ 的原因. 但是, 我们应该知道这个假设是不失一般性的. 即如果 $(\boldsymbol{\mu})_{\mathbf{0}}=1$, 那么所得到的复合过程是平凡的: 它永不移动. 另一方面, 如果 $(\boldsymbol{\mu})_{\mathbf{0}}\in(0,1)$, 那么我们可以用 $\bar{\boldsymbol{\mu}}$ 代替 $\boldsymbol{\mu}$, 其中当 $\boldsymbol{k}\neq\mathbf{0}$ 时, $(\bar{\boldsymbol{\mu}})_{\mathbf{0}}=0$ 且 $(\bar{\boldsymbol{\mu}})_{\boldsymbol{k}}=(1-(\boldsymbol{\mu})_{\mathbf{0}})^{-1}(\boldsymbol{\mu})_{\boldsymbol{k}}$, 又用 $\bar{R}=(1-(\boldsymbol{\mu})_{\mathbf{0}})R$ 代替 R. 相应于 $\bar{\boldsymbol{\mu}}$ 和 $\bar{R}$ 的复合 Poisson 过程和相应于 $\boldsymbol{\mu}$ 以及 R 的复合 Poisson 过程具有相同的分布.

所以 (4.16) 式得证.

最后来计算 $\boldsymbol{X}(t)$ 的分布. 我们知道, n 个独立同分布随机变量之和的分布是它们的分布的 n 重卷积. 因此, $\mathrm{P}(\boldsymbol{X}_n=\boldsymbol{k})=(\boldsymbol{\mu}^{\star n})_{\boldsymbol{k}}$, 其中 $(\boldsymbol{\mu}^{\star 0})_{\boldsymbol{k}}=\delta_{\mathbf{0},\boldsymbol{k}}$ 是在原点 $\mathbf{0}$ 的质量 (概率), 且对 $n\geqslant 1$ 有

$$(\boldsymbol{\mu}^{\star n})_{\boldsymbol{k}}=\sum_{\boldsymbol{j}\in\mathbb{Z}^d}(\boldsymbol{\mu}^{\star(n-1)})_{\boldsymbol{k}-\boldsymbol{j}}(\boldsymbol{\mu})_{\boldsymbol{j}}.$$

所以,

$$\mathrm{P}(\boldsymbol{X}(s+t)=\boldsymbol{k})=\sum_{n=0}^{\infty}\mathrm{P}(\boldsymbol{X}_n=\boldsymbol{k},N(Rt)=n)=e^{-Rt}\sum_{n=0}^{\infty}\frac{(Rt)^n}{n!}(\boldsymbol{\mu}^{\star n})_{\boldsymbol{k}}.$$

把上述结果和 (4.1.6) 式结合起来, 我们会发现, 对 $A\in\sigma(\{N(\tau):\tau\in[0,s]\})$,

$$\begin{aligned}
&\mathrm{P}(\{\boldsymbol{X}(s+t)=\boldsymbol{k}\}\bigcap A)\\
=&\sum_{\boldsymbol{j}\in\mathbb{Z}^d}\mathrm{P}(\{\boldsymbol{X}(s+t)=\boldsymbol{k}\}\bigcap A\bigcap\{\boldsymbol{X}(s)=\boldsymbol{j}\})\\
=&\sum_{\boldsymbol{j}\in\mathbb{Z}^d}\mathrm{P}(\{\boldsymbol{X}(s+t)-\boldsymbol{X}(s)=\boldsymbol{k}-\boldsymbol{j}\}\bigcap A\bigcap\{\boldsymbol{X}(s)=\boldsymbol{j}\})\\
=&\sum_{\boldsymbol{j}\in\mathbb{Z}^d}(\boldsymbol{P}(t))_{\boldsymbol{j}\boldsymbol{k}}\mathrm{P}(A\bigcap\{\boldsymbol{X}(s)=\boldsymbol{j}\})=\mathrm{E}[(\boldsymbol{P}(t))_{\boldsymbol{X}(s)\boldsymbol{k}},A],
\end{aligned}$$

其中

$$(\boldsymbol{P}(t))_{\boldsymbol{k}\boldsymbol{\ell}}\equiv e^{-Rt}\sum_{m=0}^{\infty}\frac{(Rt)^m}{m!}(\boldsymbol{\mu}^{\star m})_{\boldsymbol{\ell}-\boldsymbol{k}}. \tag{4.1.7}$$

这就是说, 我们已经证明了 $\{\boldsymbol{X}(t):t\geqslant 0\}$ 是一个**转移概率为** $t\rightsquigarrow\boldsymbol{P}(t)$ **的连续时间** *Markov* **过程**, 也即

$$\mathrm{P}(\boldsymbol{X}(s+t)=\boldsymbol{k}|\boldsymbol{X}(\sigma),\sigma\in[0,s])=(\boldsymbol{P}(t))_{\boldsymbol{X}(s)\boldsymbol{k}}. \tag{4.1.8}$$

作为 (4.1.8) 式的一个推论, 我们发现 $\{\boldsymbol{P}(t):t\geqslant 0\}$ 是一个**半群**. 也就是说, 它满足 Chapman-Kolmogorov **方程**

$$\boldsymbol{P}(s+t)=\boldsymbol{P}(s)\boldsymbol{P}(t),\quad s,t\in[0,\infty). \tag{4.1.9}$$

事实上,

$$\begin{aligned}
(\boldsymbol{P}(s+t))_{\mathbf{0}\boldsymbol{k}}&=\sum_{\boldsymbol{j}\in\mathbb{Z}^d}\mathrm{P}(\boldsymbol{X}(s+t)=\boldsymbol{k},\boldsymbol{X}(s)=\boldsymbol{j})\\
&=\sum_{\boldsymbol{j}\in\mathbb{Z}^d}(\boldsymbol{P}(t))_{\boldsymbol{j}\boldsymbol{\ell}}(\boldsymbol{P}(s))_{\mathbf{0}\boldsymbol{j}}=(\boldsymbol{P}(s)\boldsymbol{P}(t))_{\mathbf{0}\boldsymbol{k}},
\end{aligned}$$

如果再注意到 $(\boldsymbol{P}(\tau))_{k\ell} = (\boldsymbol{P}(\tau))_{0(\ell-k)}$, 待证的矩阵等式就直接可从上式推出.

4.2 带有界速率的 Markov 过程

在不改变 Markov 性质的前提下, 我们可以从两方面来推广上节的内容. 一种途径是使跳跃的分布与过程在跳跃时所处的位置有关. 这个改变可来自于用一个一般的 Markov 链代替复合 Poisson 过程中的随机游动. 另一种推广是增加随机性, 使得跳跃的速率与过程在跳跃前所处的位置有关. 也就是说, 在一个特定状态的停留时间与这个状态有关, 而不是所有停留时间都具有相同的均值.

4.2.1. 基本结构: 令 $\mathbb{S}$ 是一个可数状态空间, $\boldsymbol{P}$ 是一个具有性质 $(\boldsymbol{P})_{ii} = 0, i \in \mathbb{S}$ 的转移概率矩阵[3]. 再令 $\mathcal{R} \equiv \{R_i : i \in \mathbb{S}\} \subseteq [0,\infty)$ 是一族*速率*. 那么, $\mathbb{S}$ 上具有速率 $\mathcal{R}$ 和转移概率矩阵 $\boldsymbol{P}$ 的连续时间 Markov 过程是一族具有下列性质的 $\mathbb{S}$ 值随机变量 $\{X(t) : t \geqslant 0\}$:

(a) $t \rightsquigarrow X(t)$ 逐段为常数且右连续,

(b) 如果 $J_0 \equiv 0$ 并且对 $n \geqslant 1$, J_n 表示 $t \rightsquigarrow X(t)$ 的第 n 次跳跃的时间, 那么在 $\{J_{n-1} < \infty\}$ 上,

$$\begin{aligned} &\mathrm{P}(J_n > J_{n-1} + t,\ X(J_n) = j | X(\tau), \tau \in [0, J_n)) \\ &= e^{-tR_{X(J_{n-1})}}(\boldsymbol{P})_{X(J_{n-1})j}. \end{aligned} \tag{4.2.1}$$

我们首先说明, 以上性质和初始分布一起就完全决定了 $\{X(t) : t \geqslant 0\}$ 的分布, 并且我们限于考虑速率 $\mathcal{R}$ *是有界的*, 即 $\sup_i R_i < \infty$ 的情况. 其理由后面很快会明白. 事实上, 此刻我们也假设速率 $\mathcal{R}$ *是非退化的*, 即 $\mathcal{R} \subseteq (0,\infty)$. 由于非退化意味着对每个 $n \geqslant 0, \mathrm{P}(J_n < \infty) = 1$, 所以我们可以 (参看 (6.1.5)) 假设对每个 $n \geqslant 0, J_n < \infty$. 对 $n \geqslant 1$, 令 $X_n = X(J_n)$ 和 $E_n = R_{X_{n-1}}(J_n - J_{n-1})$. 根据 (4.2.1) 中的 (b), 我们有

$$\mathrm{P}(E_n > t,\ X_n = j | \{E_1, \cdots, E_{n-1}\} \bigcup \{X_0, \cdots, X_{n-1}\}) = e^{-t}(\boldsymbol{P})_{X_n j}.$$

所以, $\{X_n : n \geqslant 0\}$ 是一个转移概率矩阵为 $\boldsymbol{P}$ 的 Markov 链, 其初始分布与 $X(0)$ 的分布相同; $\{E_n : n \geqslant 1\}$ 是一列相互独立的单位指

[3] 此处作这个假定的原因和 §4.1.2 里假定 $(\boldsymbol{\mu})_0 = 0$ 是相同的, 并且此处的假定也是不失一般性的.

数随机变量, 而 $\sigma(\{X_n : n \geqslant 0\})$ 和 $\sigma(\{E_n : n \geqslant 1\})$ 相互独立. 因此, $\{X_n : n \geqslant 0\}$ 和 $\{E_n : n \geqslant 1\}$ 的联合分布是被唯一确定的. 此外, $\{X(t) : t \geqslant 0\}$ 也能从 $\{X_n : n \geqslant 0\} \bigcup \{E_n : n \geqslant 1\}$ 中得出. 也就是说, 给定 $(e_1, \cdots, e_n, \cdots) \in (0, \infty)^{\mathbb{Z}^+}$ 和 $(j_0, \cdots, j_n, \cdots) \in \mathbb{S}^{\mathbb{N}}$, 对 $\xi_n \leqslant t < \xi_{n+1}$, 定义

$$\Phi^{\mathcal{R}}(t; (e_1, \cdots, e_n, \cdots), (j_0, \cdots, j_n, \cdots)) = j_n, \tag{4.2.2}$$

其中 $\xi_0 = 0$, 当 $n \geqslant 1$ 时 $\xi_n = \sum\limits_{m=1}^{n} R_{j_m - 1}^{-1} e_m$. 显然, 对 $0 \leqslant t < \sum\limits_{m=1}^{\infty} R_{j_m - 1}^{-1} E_m$, 有

$$X(t) = \Phi^{\mathcal{R}}(t; (E_1, \cdots, E_n, \cdots), (X_0, \cdots, X_n, \cdots)). \tag{4.2.3}$$

这样, 一旦验证了 $\sum\limits_{m=1}^{\infty} R_{X_{m-1}}^{-1} E_n = \infty$ 以概率 1 成立, 那么 $\{X(t) : t \geqslant 0\}$ 的分布就是唯一确定的. 这其实就是我们假定 $\mathcal{R}$ 有界的原因. 也就是说, 不论是根据强大数律还是精确的计算 (比如说求 $\mathrm{E}[\exp(-\sum\limits_{1}^{\infty} E_m)]$), 我们都可以得到 $\sum\limits_{1}^{\infty} E_m = \infty$ 以概率 1 成立. 因此, $\mathcal{R}$ 的有界性足以保证 $\sum\limits_{1}^{\infty} R_{X_{m-1}}^{-1} E_m = \infty$ 以概率 1 成立.

为讨论[4] 退化情形, 也就是有某些 R_i 为 0 的情形, 我们令 $\mathbb{S}_0 = \{i : R_i = 0\}$ 并且确定 $\bar{\mathcal{R}}$, 使得如果 $i \notin \mathbb{S}_0$, 那么 $\bar{R}_i = R_i$, 如果 $i \in \mathbb{S}_0$, 那么 $\bar{R}_i = 1$. 我们的目标就是证明 $\{X(t) : t \geqslant 0\}$ 和 $\{\bar{X}(t \wedge \varsigma) : t \geqslant 0\}$ 同分布, 其中 $\{\bar{X}(t) : t \geqslant 0\}$ 是一个具有速率 $\bar{\mathcal{R}}$ 和转移概率矩阵 $\boldsymbol{P}$ 的过程, 而 $\varsigma \equiv \inf\{t \geqslant 0 : \bar{X}(t) \in \mathbb{S}_0\}$ 是 $t \rightsquigarrow X(t)$ 首次到达 $\mathbb{S}_0$ 的时刻. 也即, 过程 $\{X(t) : t \geqslant 0\}$ 的分布就是当到达 $\mathbb{S}_0$ 时停止的过程 $\{\bar{X}(t) : t \geqslant 0\}$ 的分布.

为了验证上面的结论, 令 $\{X_n^{(i)} : n \geqslant 0, \quad i \in \mathbb{S}_0\}$ 是一族 $\mathbb{S}$ 值随机变量, $\{\bar{E}_n : n \geqslant 1\}$ 是一族 $(0, \infty)$ 值的随机变量, 具有下列性质:

(1) $\sigma(\{X_n^{(i)} : n \geqslant 0, \quad i \in \mathbb{S}_0\})$, $\sigma(\{\bar{E}_n : n \geqslant 1\})$ 以及 $\sigma(\{X(t) : t \geqslant 0\})$ 相互独立,

(2) 对每个 $i \in \mathbb{S}_0$, $\{X_n^{(i)} : n \geqslant 0\}$ 是从 i 开始的具有转移概率矩阵 $\boldsymbol{P}$ 的 Markov 链,

[4] 下面的讨论有一些技巧性, 不需要完全掌握.

(3) $\{\bar{E}_n : n \geqslant 1\}$ 是相互独立的单位指数随机变量.
那么, 对每个 $i \in \mathbb{S}$, 由

$$\bar{X}^{(i)}(t) = \Phi^{\bar{\mathcal{R}}}(t; (\bar{E}_1, \cdots, \bar{E}_n, \cdots), (X_0^{(i)}, \cdots, X_n^{(i)}, \cdots))$$

定义了一个从 i 开始的具有速率 $\bar{\mathcal{R}}$ 和转移概率矩阵 $\boldsymbol{P}$ 的过程 $\{\bar{X}^{(i)}(t) : t \geqslant 0\}$. 最后, 定义 $\{\bar{X}(t) : t \geqslant 0\}$, 使得

$$\bar{X}(t) = \begin{cases} X(t), & \text{若 } t < \varsigma; \\ \bar{X}^{(X(\varsigma))}(t - \varsigma), & \text{若 } t \geqslant \varsigma. \end{cases}$$

显然, $X(t) = \bar{X}(t \wedge \varsigma)$. 因此, 我们只要证明 $\{\bar{X}(t) : t \geqslant 0\}$ 的分布与一个速率为 $\bar{\mathcal{R}}$ 且转移概率矩阵为 $\boldsymbol{P}$ 的过程的分布相同即可. 为此, 我们令 $\bar{J}_0 \equiv 0$, 对 $m \geqslant 1$, 令 $\bar{J}_m$ 为 $t \rightsquigarrow X(t)$ 的第 m 次跳跃的时间. 假定 $A \in \sigma(\{\bar{X}(\tau) : \tau \in [0, J_n)\})$. 我们要证明

$$\mathrm{P}(\{\bar{J}_n > \bar{J}_{n-1} + t, \bar{X}(\bar{J}_n) = j\} \bigcap A) = \mathrm{E}\left[e^{-t\bar{R}_{\bar{X}(\bar{J}_{n-1})}}(\boldsymbol{P})_{\bar{X}(\bar{J}_{n-1})j}, A\right]. \tag{$*$}$$

由于我们总能把 A 写成形式为 $\{\bar{X}(\bar{J}_m) = j_m, 0 \leqslant m < n\}$ 的互不相交的集合的并, 所以我们不妨假定 A 本身就具有这种形式. 如果对每个 $0 \leqslant m < n$, $R_{j_m} > 0$, 那么 $A \in \sigma(\{X(\sigma) : \sigma \in [0, J_n)\})$, 并且在 A 上, 对 $0 \leqslant \ell \leqslant n$ 有 $(\bar{J}_\ell, \bar{X}(\bar{J}_\ell)) = (J_\ell, X(J_\ell))$. 因此在这种情况下, $(*)$ 式成立. 如果对某个 $0 \leqslant \ell < n$ 有 $j_\ell \in \mathbb{S}_0$, 用 m 记第一个这样的 ℓ, 令 $i = j_m$, 并令 $\{\bar{J}_\ell^{(i)} : \ell \geqslant 0\}$ 为 $t \rightsquigarrow \bar{X}^{(i)}(t)$ 的跳跃时间. 那么, 可写 $A = B \bigcap C$, 其中 $B = \{X(J_\ell) = j_\ell, \quad 0 \leqslant \ell \leqslant m\}$,

$$C = \begin{cases} \{\bar{X}^{(i)}(\bar{J}_{\ell-m}^{(i)}) = j_\ell, \quad m < \ell < n-1\}, & \text{若 } 0 \leqslant m < n-1; \\ \Omega, & \text{若 } m = n-1. \end{cases}$$

另外, 在 A 上, 对 $\ell > m$ 有 $(\bar{J}_n, \bar{X}(\bar{J}_\ell)) = (\bar{J}_{\ell-m}^{(i)}, \bar{X}^{(i)}(\bar{J}_{\ell-m}^{(i)}))$. 从而,

$$\begin{aligned}
&\mathrm{P}\left(\{\bar{J}_n > \bar{J}_{n-1} + t, \bar{X}(\bar{J}_n) = j\} \bigcap A\right) \\
&= \mathrm{P}\left(\{\bar{J}_{n-m}^{(i)} > \bar{J}_{n-m-1}^{(i)} + t, \bar{X}^{(i)}(\bar{J}_{n-m}^{(i)}) = j\} \bigcap C\right) \mathrm{P}(B) \\
&= \mathrm{E}\left[\exp(-t\bar{R}_{\bar{X}^{(i)}(\bar{J}_{n-m-1}^{(i)})})(\boldsymbol{P})_{\bar{X}^{(i)}(\bar{J}_{n-m-1}^{(i)})j}, C\right] \mathrm{P}(B) \\
&= \mathrm{E}\left[e^{-t\bar{R}_{\bar{X}(\bar{J}_{n-1})}}(\boldsymbol{P})_{\bar{X}(\bar{J}_{n-1})j}, A\right],
\end{aligned}$$

因此, 在这种情况下 (*) 式也成立.

鉴于前面所述, 我们知道 $\{X(t): t \geqslant 0\}$ 的分布由 (4.2.1) 式和它的初始分布唯一确定. 事实上, 它和由 (4.2.3) 式确定的过程同分布. 当然, 在退化情形下, 即使 (4.2.3) 式成立, 指数随机变量 $\{E_n : n \geqslant 1\}$ 和 Markov 链 $\{X_n : n \geqslant 0\}$ 都不是过程 $\{X(t): t \geqslant 0\}$ 的函数. 但是, 既然我们已经证明了它们的分布的唯一性, 就不需要再去考虑它了, 所以我们总是可以假定我们的过程是由 (4.2.3) 式给定的.

4.2.2. Markov 性: 我们仍假设速率是有界的, 现在要验证上面所述的过程 $\{X(t): t \geqslant 0\}$ 具有 Markov 性:

$$\mathrm{P}(X(s+t)=j|X(\tau), \tau \in [0,s]) = (\boldsymbol{P}(t))_{X(s)j}, \tag{4.2.4}$$

其中 $(\boldsymbol{P}(t))_{ij} \equiv \mathrm{P}(X(t)=j|X(0)=i)$. 为了证明它, 下面的结论是有用的:

$$\begin{aligned}
&\xi_m \leqslant s < \xi_{m+1} \\
&\Longrightarrow \Phi^{\mathcal{R}}(s+t;(e_1,\cdots,e_n,\cdots),(j_0,\cdots,j_n,\cdots)) \\
&= \ \Phi^{\mathcal{R}}(t;(e_{m+1}-R_{j_m}(s-\xi_m),e_{m+2},\cdots,e_{m+n},\cdots), \\
&\quad\ (j_m,\cdots,j_{m+n},\cdots)).
\end{aligned} \tag{4.2.5}$$

给定 $A \in \sigma(\{X(\tau) : \tau \in [0,s]\})$, 并且假设在 A 上, $X(s)=i$. 我们所要做的就是验证

$$\mathrm{P}(\{X(s+t)=j\}\bigcap A) = (\boldsymbol{P}(t))_{ij}\mathrm{P}(A). \tag{*}$$

为此, 我们令 $A_m \equiv A\bigcap\{J_m < s \leqslant J_{m+1}\} = \{E_{m+1} > R_i(s-J_m)\}\bigcap B_m$, 其中 $\{J_m \leqslant s\} \supseteq B_m \in \sigma(\{E_1,\cdots,E_m\}\bigcup\{X_0,\cdots,X_m\})$. 那么

$$\begin{aligned}
&\mathrm{P}(\{X(s+t)=j\}\bigcap A) \\
&= \sum_{m=0}^{\infty} \mathrm{P}(\{X(s+t)=j\}\bigcap A_m) \\
&= \sum_{m=0}^{\infty} \mathrm{P}(\{X(s+t)=j, \quad E_{m+1} > R_i(s-J_m)\}\bigcap B_m).
\end{aligned}$$

根据 (4.2.3), (4.2.5) 和 (4.1.3) 式, 有

$$\begin{aligned}
&\mathrm{P}(\{X(s+t)=j, E_{m+1}>R_i(s-J_m)\}\bigcap B_m)\\
&=\mathrm{P}(\{\Phi^{\mathcal{R}}(t;(E_{m+1}-R_i(s-J_m)),E_{m+2},\cdots,E_{m+n},\cdots),\\
&\quad(i,X_{m+1},\cdots,X_{m+n},\cdots))=j\}\bigcap\{E_{m+1}>R_i(s-J_m)\}\bigcap B_m)\\
&=\mathrm{P}(X(t)=j|X(0)=i)\mathrm{E}[e^{-R_i(s-J_m)},B_m]=(\boldsymbol{P}(t))_{ij}\mathrm{P}(A_m),
\end{aligned}$$

由此很容易就推出 (*) 式.

就像 (4.1.8) 式能推出转移概率矩阵的半群性质一样 (参看 (4.1.9)), (4.2.4) 式也能推出 $\{\boldsymbol{P}(t):t\geqslant 0\}$ 是一个半群. 实际上, 证明方法是一样的:

$$\begin{aligned}
(\boldsymbol{P}(t))_{ij}&=\sum_{k\in\mathbb{S}}\mathrm{P}(X(s+t)=j,X(s)=k|X(0)=i)\\
&=\sum_{k\in\mathbb{S}}(\boldsymbol{P}(t))_{kj}\mathrm{P}(X(s)=k)=\sum_{k\in\mathbb{S}}(\boldsymbol{P}(t))_{kj}(\boldsymbol{P}(s))_{ik}\\
&=(\boldsymbol{P}(s)\boldsymbol{P}(t))_{ij}.
\end{aligned}$$

此外, 重要的一点 (至少在原则上) 是认识到*Markov*过程的分布由其初始分布和转移概率矩阵 $\{\boldsymbol{P}(t):t>0\}$ 的半群唯一决定. 准确地说, 假定 $\{X(t):t\geqslant 0\}$ 是一族使得 (4.2.4) 式成立的 $\mathbb{S}$ 值随机变量, 并令 $\boldsymbol{\mu}$ 是它的初始分布 (也即 $X(0)$ 的分布). 那么, 对所有 $n\geqslant 1$, $0=t_0<t_1<\cdots<t_n$, 以及 $j_0,\cdots,j_n\in\mathbb{S}$,

$$\begin{aligned}
&\mathrm{P}(X(t_m)=j_m,0\leqslant m\leqslant n)\\
&=(\boldsymbol{\mu})_{j_0}(\boldsymbol{P}(t_1-t_0))_{j_0j_1}\cdots(\boldsymbol{P}(t_n-t_{n-1}))_{j_{n-1}j_n}.
\end{aligned}\tag{4.2.6}$$

为了验证它, 我们首先指出, 由 (4.2.4) 式,

$$\mathrm{P}(X(t_0)=j_0,\ X(t_1)=j_1)=(\boldsymbol{P}(t_1))_{j_0j_1}(\boldsymbol{\mu})_{j_0}=(\boldsymbol{\mu})_{j_0}(\boldsymbol{P}(t_1-t_0))_{j_0j_1}.$$

因此, 对 $n=1$, (4.2.6) 式成立. 现在给定 $n\geqslant 2$, 假设 (4.2.6) 对 $n-1$ 成立, 记 $A=\{X(t_m)=j_m:0\leqslant m\leqslant n-1\}$. 那么, 根据 (4.2.4) 式, 有

$$\begin{aligned}
\mathrm{P}(X(t_m)=j_m,\ 0\leqslant m\leqslant n)&=\mathrm{P}(\{X(t_n)-j_n\}\bigcap A)\\
&=(\boldsymbol{P}(t_n-t_{n-1}))_{j_{n-1}j_n}\mathrm{P}(A),
\end{aligned}$$

因而由归纳假设可得 (4.2.6) 式成立. 最后, 根据 §6.1.4 中的定理 6.1.6, $\{X(t):t\geqslant 0\}$ 的分布完全由形式为 $\{X(t_m)=j_m,0\leqslant m\leqslant n\}$ 的集合的概率决定, 从而就证明了唯一性.

4.2.3. Q-矩阵和 Kolmogorov 向后方程: 正如我们在前一节末尾所述, 除了它的初始分布, 一个 Markov 过程的分布完全由半群 $\{\boldsymbol{P}(t): t \geqslant 0\}$ 决定. 因此, 直接从包含有速率 $\mathcal{R}$ 和转移概率 $\boldsymbol{P}$ 的数据中计算出转移概率 $\boldsymbol{P}(t)$ 的方法就显得很重要了.

基于处理实值函数的经验, 人们会认为 (4.1.9) 式意味着 $\boldsymbol{P}(t)$ 可以对某个 $\boldsymbol{Q}$ 表示成 $e^{t\boldsymbol{Q}}$. 事实上, $\boldsymbol{Q}$ 应该可以通过在 $t=0$ 处对 $t \rightsquigarrow P(t)$ 微分得到.

为了证实上述推测, 首先, 我们来证明

$$(\boldsymbol{P}(t))_{ij} = \delta_{i,j} e^{-tR_i} + R_i \int_0^t e^{-\tau R_i} (\boldsymbol{P}\boldsymbol{P}(t-\tau))_{ij} d\tau. \qquad (*)$$

$R_i = 0$ 的情形是显然的, 所以我们假设 $R_i > 0$. 由于

$$(\boldsymbol{P}(t))_{ij} = \delta_{ij} \mathrm{P}(E_1 > tR_i | X(0) = i) + \mathrm{P}(E_1 \leqslant tR_i, X(t) = j | X(0) = i),$$

且 (参看 (4.2.3) 和 (4.2.5))

$$\begin{aligned}
&\mathrm{P}(E_1 \leqslant tR_i, X(t) = j | X(0) = i) \\
=& \mathrm{P}(\Phi^{\mathcal{R},\boldsymbol{P}}(t - R_i^{-1}E_1; (E_2, \cdots, E_n, \cdots), (X_1, \cdots, X_n, \cdots)) = j, \\
& E_1 \leqslant tR_i | X_0 = i) \\
=& \mathrm{E}[(\boldsymbol{P}(t - R_i^{-1}E_1))_{X_1 j}, E_1 \leqslant R_i t | X_0 = i] \\
=& R_i \int_0^t e^{-\tau R_i} \sum_{k \in \mathbb{S}} (\boldsymbol{P})_{ik} (\boldsymbol{P}(t-\tau))_{kj} d\tau.
\end{aligned}$$

这样就证明了 $(*)$ 式.

$(*)$ 式是 Kolmogorov 的一个著名方程的积分形式. 如果对 $(*)$ 式关于 t 微分, 就可得到 *Kolmogorov 向后方程*:

$$\frac{d}{dt}\boldsymbol{P}(t)_{ij} = -R_i \boldsymbol{P}(t)_{ij} + R_i (\boldsymbol{P}\boldsymbol{P}(t))_{ij},$$

写成矩阵形式就是

$$\frac{d}{dt}\boldsymbol{P}(t) = \boldsymbol{Q}\boldsymbol{P}(t), \quad \boldsymbol{P}(0) = \boldsymbol{I}, \quad \text{其中 } \boldsymbol{Q} = \boldsymbol{R}(\boldsymbol{P} - \boldsymbol{I}), \qquad (4.2.7)$$

这里的 $\boldsymbol{R}$ 是第 i 个对角元为 R_i 的对角矩阵. 之所以在前面加形容词 "向后" 是因为 (4.2.7) 式描述了作为时间 t 和称为*向后变量* i 的函数 $t \rightsquigarrow (\boldsymbol{P}(t))_{ij}$ 的变化. "向后" 一词出于如下的背景: 某人沿着路径

$t \rightsquigarrow X(t)$ 旅行, 则他出发的地点 i 是当他向后回望时所看到的变量. 不过, 矩阵 $\boldsymbol{Q}$ 的术语缺少一定的启发性. 概率学家把任何一个非对角线元素非负、且行和为零的矩阵都称为 Q-矩阵.

4.2.4. Kolmogorov 向前方程: 在 §3.2.1 开始时曾介绍过范数 $\|\cdot\|_{u,v}$. 从 (4.2.7) 式, 我们有

$$\boldsymbol{P}(t) = \boldsymbol{I} + \int_0^t \boldsymbol{Q}\boldsymbol{P}(\tau)\,d\tau = \boldsymbol{I} + t\boldsymbol{Q} + \int_0^t (t-\tau)\boldsymbol{Q}^2\boldsymbol{P}(\tau)\,d\tau,$$

因此

$$\|\boldsymbol{P}(t) - \boldsymbol{I} - t\boldsymbol{Q}\|_{u,v} \leqslant \frac{\|\boldsymbol{Q}\|_{u,v}^2 t^2}{2}. \tag{4.2.8}$$

因为由半群的性质 (参看 (4.1.9)),

$$\boldsymbol{P}(t+h) - \boldsymbol{P}(t) - h\boldsymbol{Q}\boldsymbol{P}(t) = \boldsymbol{P}(t)(\boldsymbol{P}(h) - \boldsymbol{I} - h\boldsymbol{Q}),$$

从而可以得到所谓的 *Kolmogorov 向前方程*

$$\frac{d}{dt}\boldsymbol{P}(t) = \boldsymbol{P}(t)\boldsymbol{Q}, \quad \boldsymbol{P}(0) = \boldsymbol{I}, \tag{4.2.9}$$

之所以这样叫是因为它描述了 $t \rightsquigarrow \boldsymbol{P}(t)$ 作为 *向前变量* j 的函数的变化, 这里 j 是当旅行者 "向前展望" 时所看到的变量.

4.2.5. 解 Kolmogorov 方程: 正如以前提到的, (4.1.9) 式指出 $\boldsymbol{P}(t) = e^{t\dot{\boldsymbol{P}}(0)}$, 并且根据 (4.2.7) 式, 我们有 $\dot{\boldsymbol{P}}(0) = \boldsymbol{Q}$. 因此, 我们猜测 $\boldsymbol{P}(t) = e^{t\boldsymbol{Q}}$, 这里指数的意义由幂级数给出:

$$e^{\boldsymbol{M}} \equiv \sum_{m=0}^{\infty} \frac{\boldsymbol{M}^m}{m!}, \quad \boldsymbol{M} \in M_{u,v}(\mathbb{S}). \tag{4.2.10}$$

因为 $\|\boldsymbol{M}^m\|_{u,v} \leqslant \|\boldsymbol{M}\|_{u,v}^m$, 毫无疑问对每一个 $\boldsymbol{M} \in M_{u,v}(\mathbb{S})$ 上述级数收敛. 事实上, 因为

$$\left\| e^{\boldsymbol{M}} - \sum_{m=0}^{n-1} \frac{\boldsymbol{M}^m}{m!} \right\|_{u,v} \leqslant \sum_{m=n}^{\infty} \frac{\|\boldsymbol{M}\|_{u,v}^m}{m!},$$

所以

$$\left\| e^{\boldsymbol{M}} - \sum_{m=0}^{n-1} \frac{\boldsymbol{M}^m}{m!} \right\|_{u,v} \leqslant \frac{\|\boldsymbol{M}\|_{u,v}^n}{n!} e^{\|\boldsymbol{M}\|_{u,v}}. \tag{4.2.11}$$

也容易验证 (参看后面的习题 4.5.2): 对可交换的 $\boldsymbol{M}_1$, $\boldsymbol{M}_2 \in M_{u,v}(\mathbb{S})$,

$$e^{\boldsymbol{M}_1 + \boldsymbol{M}_2} = e^{\boldsymbol{M}_1} e^{\boldsymbol{M}_2}. \tag{4.2.12}$$

特别地, 这意味着 $t \rightsquigarrow e^{t\boldsymbol{Q}}$ 具有半群性: $e^{(s+t)\boldsymbol{Q}} = e^{s\boldsymbol{Q}}e^{t\boldsymbol{Q}}$. 最后, 由于

$$\left\|e^{h\boldsymbol{Q}} - \boldsymbol{I} - h\boldsymbol{Q}\right\|_{u,v} \leqslant \frac{\|\boldsymbol{Q}\|_{u,v}^2 h^2}{2} e^{h\|\boldsymbol{Q}\|_{u,v}},$$

我们得到

$$\begin{aligned}\left\|\frac{e^{(t+h)\boldsymbol{Q}} - e^{t\boldsymbol{Q}}}{h} - \boldsymbol{Q}e^{t\boldsymbol{Q}}\right\|_{u,v} &= \left\|e^{t\boldsymbol{Q}}\left(\frac{e^{h\boldsymbol{Q}} - \boldsymbol{I}}{h} - \boldsymbol{Q}\right)\right\|_{u,v} \\ &\leqslant \frac{\|\boldsymbol{Q}\|_{u,v}^2 h e^{(t+h)\|\boldsymbol{Q}\|_{u,v}}}{2}.\end{aligned}$$

因此 $\frac{d}{dt}e^{t\boldsymbol{Q}} = e^{t\boldsymbol{Q}}\boldsymbol{Q}$.

有了这些准备, 容易完成等式

$$\boldsymbol{P}(t) = e^{t\boldsymbol{Q}} \equiv \sum_{m=0}^{\infty} \frac{t^m \boldsymbol{Q}^m}{m!}, \quad t \in [0, \infty) \tag{4.2.13}$$

的证明了. 事实上, 由前式和 (4.2.7) 式, 对任意 $t > 0$ 和 $\tau \in [0, t]$, 有

$$\frac{d}{d\tau} e^{(t-\tau)\boldsymbol{Q}} \boldsymbol{P}(\tau) = e^{(t-\tau)\boldsymbol{Q}} (\boldsymbol{Q} - \boldsymbol{Q}) \boldsymbol{P}(\tau) = \boldsymbol{0},$$

因此 $\tau \in [0, t] \mapsto e^{(t-\tau)\boldsymbol{Q}}\boldsymbol{P}(\tau) \in M_{u,v}(\mathbb{S})$ 为常数.

在结束讨论之前, 还应指出重要的一点. 通过 (4.2.13) 式, 我们已经证明了: 对每一个 $t > 0$, $e^{t\boldsymbol{Q}}$ 是一个转移概率矩阵, 但是这个事实从 $e^{t\boldsymbol{Q}}$ 的幂级数展开式看并不明显. 特别地, 如果没有进一步的考虑, 很难明白为什么 $e^{t\boldsymbol{Q}}$ 的元素是非负的, 为什么 $\|e^{t\boldsymbol{Q}}\|_{u,v} = 1$, 而与 $\|\boldsymbol{Q}\|_{u,v}$ 无关. 因此, 我们将提供另一途径来理解这些性质, 这并不需要把 $e^{t\boldsymbol{Q}}$ 视为 $\boldsymbol{P}(t)$, 而是令

$$\hat{\boldsymbol{P}}_{ij} = \begin{cases} \dfrac{Q_{ij}}{M}, & 若\ i \neq j; \\ 1 + \dfrac{Q_{ij}}{M}, & 若\ i = j. \end{cases} \quad 其中\ M \equiv \sup_i R_i = \sup_i(-(\boldsymbol{Q}_{ii})).$$

则 $\hat{\boldsymbol{P}}$ 是一个转移概率矩阵, 且 $\boldsymbol{Q} = M(\hat{\boldsymbol{P}} - \boldsymbol{I})$. 因此, 由于 $\boldsymbol{I}$ 与 $\hat{\boldsymbol{P}}$ 可交换, 所以

$$e^{t\boldsymbol{Q}} = e^{-tM\boldsymbol{I}} e^{tM\hat{\boldsymbol{P}}} = e^{-tM} \sum_{m=0}^{\infty} \frac{(tM)^m}{m!} \hat{\boldsymbol{P}}^m.$$

从这个表示式容易看出，$e^{t\boldsymbol{Q}}$ 的元素是非负的, 而且

$$\sum_{j\in\mathbb{S}}(e^{t\boldsymbol{Q}})_{ij}=e^{-tM}\sum_{m=1}^{\infty}\frac{(tM)^m}{m!}\left(\sum_{j\in\mathbb{S}}(\hat{\boldsymbol{P}}^m)_{ij}\right)=1.$$

欣赏这一推导时, 最好注意到这都是因为 $t\geqslant 0$ 的缘故; 当 $t<0$ 时, $e^{t\boldsymbol{Q}}$ 的元素不一定是非负的, 而且 $\|e^{t\boldsymbol{Q}}\|_{u,v}$ 可能和 $e^{|t|\|\boldsymbol{Q}\|_{u,v}}$ 一样大.

4.2.6. 具有无穷小特征的 Markov 过程: 在某些应用中, 根据速率 $\mathcal{R}$ 和转移概率 $\boldsymbol{P}$ 来描述 Markov 过程 $\{X(t):t\geqslant 0\}$ 的最自然的方法是给出它的初始分布, 并且在给定直至 $t\geqslant 0$ 的过去 $\sigma(\{X(\tau):\tau\in[0,t]\})$ 的条件下, 在 $t+h$ 时刻 (其中 $h>0$ 很小) 离开 $X(t)$ 的概率近似为 $hR_{X(t)}$, 而在给定 $\sigma(\{X(\tau):\tau\in[0,t]\}\bigcup\{X(t+h)\neq X(t)\})$(即知道它的过去并知道它已经移动) 的条件下, $X(t+h)=j$ 的概率近似为 $(\boldsymbol{P})_{X(t)j}$. 在这样的描述中, $\mathcal{R}$ 和 $\boldsymbol{P}$ 变为过程的无穷小特征, 问题在于是否能够从这个描述中重新构造过程的分布. 在试图重构之前我们将使描述更为量化, 也即, 存在 $\epsilon(h):(0,\infty)\to(0,\infty)$, 且当 $h\searrow 0$ 时 ϵ 趋近于 0, 使得

$$\left|\mathrm{P}(X(t+h)\neq X(t)|X(\tau),\tau\in[0,t])-hR_{X(t)}\right|\leqslant h\epsilon(h),$$

且

$$\left|\mathrm{P}(X(t+h)=j|X(\tau),\tau\in[0,t],X(t+h)\neq X(t))-(\boldsymbol{P})_{X(t)j}\right|\leqslant\epsilon(h).$$

很明显, 上面的第一个不等式等价于

$$\left|\mathrm{P}(X(t+h)=X(t)|X(\tau),\tau\in[0,t])-1-h(\boldsymbol{Q})_{X(t)X(t)}\right|\leqslant h\epsilon(h).$$

而由这两个式子得到, 当 $j\neq X(t)$ 时,

$$\left|\mathrm{P}(X(t+h)=j|X(\tau),\tau\in[0,t])-h(\boldsymbol{Q})_{X(t)j}\right|\leqslant h\epsilon(h)(R_{X(t)}+\epsilon(h)).$$

因此, 并由于我们假定速率有界, 上述表明

$$\left|\mathrm{P}(X(t+h)=j|X(\tau),\tau\in[0,t])-\delta_{X(t),j}-h\boldsymbol{Q}_{X(t)j}\right|\leqslant h\epsilon'(h),\qquad(*)$$

这里, 当 $h\searrow 0$ 时, 也有 $\epsilon'(h)\to 0$.

我们现在说明, 利用 (*) 足以证明 $\boldsymbol{P}(t)=e^{t\boldsymbol{Q}}$ 满足 (4.2.4) 式, 因此根据 §4.2.2 接近末尾时讨论的结果, $\{X(t):t\geqslant 0\}$ 是相应于

速率 $\mathcal{R}$ 和转移概率矩阵 $\boldsymbol{P}$ 的 Markov 过程. 为此, 给定 $s \geqslant 0$ 和 $A \in \sigma(\{X(\tau) : \tau \in [0, s]\})$, 选择行向量 $\{\boldsymbol{\mu}(t); t \geqslant 0\}$ 使得对每一 $j \in \mathbb{S}$, $(\boldsymbol{\mu}(t))_j = \mathrm{P}(\{X(s+t) = j\} \bigcap A)$. 于是, 由

$$(\boldsymbol{\mu}(t+h))_j = \mathrm{E}[\mathrm{P}(X(t+h) = j | X(\tau), \tau \in [0, t]), A]$$

和 $(\boldsymbol{\mu}(t))_j = \mathrm{E}[\delta_{X(t),j}, A]$, 从 $(*)$, 对 $t \geqslant 0$ 和 $h \geqslant 0$ 我们得到

$$\begin{aligned} h\epsilon'(h) &\geqslant \left|(\boldsymbol{\mu}(t+h))_j - (\boldsymbol{\mu}(t))_j - h\mathrm{E}[(\boldsymbol{Q})_{X(s+t)j}, A]\right| \\ &= \left|(\boldsymbol{\mu}(t+h))_j - (\boldsymbol{\mu}(t))_j - h(\boldsymbol{\mu}(t)\boldsymbol{Q})_j\right|. \end{aligned}$$

因此, 当 $t > 0$ 时, $\frac{d}{dt}\boldsymbol{\mu}(t) = \boldsymbol{\mu}(t)\boldsymbol{Q}$, 进而对 $\tau \in (0, t)$,

$$\frac{d}{d\tau}\boldsymbol{\mu}(\tau)e^{(t-\tau)\boldsymbol{Q}} = (\boldsymbol{\mu}(\tau)\boldsymbol{Q} - \boldsymbol{\mu}(\tau)\boldsymbol{Q})e^{(t-\tau)\boldsymbol{Q}} = \mathbf{0}\ .$$

从上式我们得到

$$\mathrm{P}(\{X(s+t) = j\} \textstyle\bigcap A) = \boldsymbol{\mu}(t) = \boldsymbol{\mu}(0)e^{t\boldsymbol{Q}} = \boldsymbol{\mu}(0)\boldsymbol{P}(t).$$

它等价于

$$\mathrm{P}(\{X(t) = j\} \textstyle\bigcap A) = \mathrm{E}[(\boldsymbol{P}(t))_{X(s)j}, A].$$

这就是说, 我们已经证明: 从 $(*)$ 可推出 (4.2.4) 式. 因此, 根据 §4.2.2 的最后部分, 我们得到 $\{X(t) : t \geqslant 0\}$ 的分布是相应于速率 $\mathcal{R}$ 和转移概率矩阵 $\boldsymbol{P}$ 的 Markov 过程的分布.

4.3 无界速率

直到现在我们一直假定速率 $\mathcal{R} = \{R_i :\ i \in \mathbb{S}\}$ 是有界的, 在 §4.2.1 中这一假定的实质应用在于它保证了以概率 1 有 $J_n \nearrow \infty$, 从而对所有时间我们的马尔可夫过程被完全确定. 正如我们下面将看到的 (参看习题 4.5.5 和 4.5.6), 当速率无界时, $J_\infty = \lim_{n\to\infty} J_n$ 可能以正概率有限. 在这种情形从我们的描述中无法得知在时间区间 $[J_\infty, \infty)$ 过程的性状. 本节我们将给出保证 $J_\infty = \infty$ 以概率 1 成立的条件, 而不是关于 $\mathcal{R}$ 有界的条件. 另外, 我们将非常粗略地讨论: 如何处理 $J_\infty < \infty$ 以正概率成立的情形.

4.3.1. 爆炸: 这里的背景与 §4.2.1 的相同, 只是现在我们不再假定速

率是有界的. 因此, 如果 $t \rightsquigarrow X(t)$ 由 (4.2.3) 式给定, J_n 表示第 n 次跳跃的时刻 (即满足 $X(t) \neq X(t-)$ 的第 n 个 t), 且 $J_\infty \equiv \lim_{n\to\infty} J_n$, 则仅仅到时刻 J_∞ 为止, $t \rightsquigarrow X(t)$ 的分布是被唯一决定的.

第一步我们将证明以概率 1, J_∞ 与过程从状态空间爆炸出的时刻相同. 确切地说, 选取 $\mathbb{S}$ 的一列非空有限子集 $\{F_N : N \geqslant 1\}$ 作为它的穷举. 即对每一 N, F_N 是 $\mathbb{S}$ 的一个非空有限子集, $F_N \subseteq F_{N+1}$, 且 $\mathbb{S} = \bigcup_N F_N$. 其次, 取 $\mathcal{R}^{(N)}$ 为由

$$R_i^{(N)} = \begin{cases} R_i, & \text{若 } i \in F_N; \\ 0, & \text{若 } i \notin F_N \end{cases}$$

给定的速率集. 则对每一 $N \geqslant 1$, 由 $\mathcal{R}^{(N)}$ 代替 $\mathcal{R}$, (4.2.3) 式决定一个马尔可夫过程 $\{X^{(N)}(t) : t \geqslant 0\}$. 进而, 如果 $\zeta_N \equiv \inf\{t \geqslant 0 : X^{(N)}(t) \notin F_N\}$, 则对 $t \in [0, \zeta_N] \bigcap [0, \infty)$ 有 $X^{(N+1)}(t) = X^{(N)}(t)$. 因此, $\zeta_N \leqslant \zeta_{N+1}$, 且爆炸时间 $e \equiv \lim_{N\to\infty} \zeta_N$ (在 $[0, \infty]$上) 存在.

4.3.1 定理. $e = J_\infty$ 以概率 1 成立, 且对每一 $N \geqslant 1$, $t \in [0, \infty)$, 有 $X^{(N)}(t) = X(t \wedge \zeta_N)$. 特别地, 如果 $\mathrm{P}(e = \infty) = 1$, 则 (4.2.1) 式描述的性质和 $X(0)$ 的分布一起唯一决定了过程 $\{X(t) : t \geqslant 0\}$ 的分布.

证明: 因为 $\{e \neq J_\infty\}$ 可表示为当 T 取遍正有理数时集合 $\{e > T \geqslant J_\infty\} \bigcup \{J_\infty > T \geqslant e\}$ 的并, 所以根据 (6.1.5) 式, 为证明 $e = J_\infty$ 以概率 1 成立, 我们只需证明对每一 $T > 0$, 有 $\mathrm{P}(e > T \geqslant J_\infty) = 0 = \mathrm{P}(J_\infty > T \geqslant e)$. 为此, 首先假设对某一 T, $\mathrm{P}(e > T \geqslant J_\infty) > 0$. 于是存在一个 N 使得 $\mathrm{P}(\zeta_N > T \geqslant J_\infty) > 0$. 另一方面, $\zeta_N > T \geqslant J_\infty \implies T \geqslant J_\infty \geqslant r_N^{-1} \sum_1^\infty E_m$, 这里 $r_N = \sup_{j \in F_N} R_j$. 因此产生矛盾

$$0 < \mathrm{P}(\zeta_N > T \geqslant J_\infty) \leq \mathrm{P}\left(\sum_{m=1}^{\infty} E_m \leqslant r_N T\right) = 0.$$

下面假设 $\mathrm{P}(J_\infty > T \geqslant e) > 0$. 于是存在一个 $n \geqslant 1$ 使得 $\mathrm{P}(J_n > T \geqslant e) > 0$. 另一方面, 如果 $\boldsymbol{\mu}$ 是 $X(0)$ 的分布, 则当 $N \to \infty$ 时, $\sum_{m=0}^{n} \sum_{j \notin F_N} (\boldsymbol{\mu P}^m)_j \to 0$, 因此

$$\begin{aligned} \mathrm{P}(J_n > T \geqslant e) &\leqslant \mathrm{P}(J_n > T \geqslant \zeta_N) \leqslant \mathrm{P}(\exists 0 \leqslant m \leqslant n,\ X_m \notin F_N) \\ &\leqslant \sum_{m=0}^{n} \sum_{j \notin F_N} (\boldsymbol{\mu P}^m)_j \to 0, \quad N \to \infty. \end{aligned}$$

即 $\mathrm{P}(J_n > T \geqslant \boldsymbol{e}) = 0$.

在有了上面的结果以后, 易知只要 $t \in [0, \zeta_N)$, 则有 $X(t) = X^{(N)}(t)$, 且当 $\zeta_N < \infty$ 时 $X(\zeta_N) = X^{(N)}(\zeta_N)$. 因此对所有 $N \in \mathbb{N}$ 和 $t \geqslant 0$, $X(t \wedge \zeta_N) = X^{(N)}(t)$. 最后, 由于 (4.2.1) 式 (其中的 $\mathcal{R}$ 用 $\mathcal{R}^{(N)}$ 代替) 和初始分布唯一决定了 $\{X^{(N)}(t) : t \geqslant 0\}$ 的分布, 从而对每一 $N \in \mathbb{N}$, $\{X(t \wedge \zeta_N) : t \geqslant 0\}$ 的分布由初始分布和 (4.2.1) 唯一决定. 因此, 当 $\mathrm{P}(\boldsymbol{e} = \infty) = 1$ 时, $\{X(t) : t \geqslant 0\}$ 的分布也由初始分布和 (4.2.1) 式唯一决定. □

4.3.2 推论. *如果对所有* $i \in \mathbb{S}$, $\mathrm{P}(\boldsymbol{e} = \infty | X(0) = i) = 1$, *且* $(\boldsymbol{P}(t))_{ij} \equiv \mathrm{P}(X(t) = j | X(0) = i)$, *则对每一初始分布* $\boldsymbol{\mu}$ *和* $T > 0$,

$$\lim_{N \to \infty} \sup_{t \in (0,T]} \|\boldsymbol{\mu} \boldsymbol{P}^{(N)}(t) - \boldsymbol{\mu} \boldsymbol{P}(t)\|_v = 0, \tag{4.3.3}$$

其中 $\{\boldsymbol{P}^{(N)} : t > 0\}$ *是由* $\mathcal{R}^{(N)}$ *和* $\boldsymbol{P}$ *决定的半群. 进一步,* $\{X(t) : t \geqslant 0\}$ *满足 (4.2.4) 式中的 Markov 性. 最后,* $\{\boldsymbol{P}(t) : t \geqslant 0\}$ *是一个半群, 它满足 Kolmogorov 向后方程: 对每一对* $(i, j) \in \mathbb{S}^2$,

$$(\boldsymbol{P}(t))_{ij} = \delta_{i,j} + \int_0^t (\boldsymbol{Q} \boldsymbol{P}(\tau))_{ij} d\tau. \tag{4.3.4}$$

证明: 首先注意到

$$\mathrm{P}(\boldsymbol{e} = \infty) = \sum_{i \in \mathbb{S}} (\boldsymbol{\mu})_i \mathrm{P}(\boldsymbol{e} = \infty | X(0) = i) = 1,$$

因此对每一 $T \in [0, \infty)$, $\lim\limits_{N \to \infty} \mathrm{P}(\zeta_N \leqslant T) = 0$. 同时, 由定理 4.3.1, 对所有 $0 \leqslant t \leqslant T$ 和 $j \in \mathbb{S}$, 有

$$|(\boldsymbol{\mu} \boldsymbol{P}^{(N)}(t))_j - (\boldsymbol{\mu} \boldsymbol{P}(t))_j| \leqslant \mathrm{P}(X(t) = j, \zeta_N \leqslant T).$$

从而

$$\sup_{t \in (0,T]} \|\boldsymbol{\mu} \boldsymbol{P}^{(N)}(t) - \boldsymbol{\mu} \boldsymbol{P}(t)\|_v \leqslant \mathrm{P}(\zeta_N \leqslant T) \to 0, \quad N \to \infty.$$

为证明 (4.2.4) 式的 Markov 性, 给定 $A \in \sigma(\{X(\tau) : \tau \in [0, s]\})$, 且假定在 A 上 $X(s) = i$. 那么, 因为 $A \bigcap \{\zeta_N > s\} \in \sigma(\{X^{(N)}(\tau) : \tau \in [0, s]\})$, 由 $\{X^{(N)}(t) : t \geqslant 0\}$ 的 Markov 性和定理 4.3.1 的结果可知

$$\begin{aligned}
&\mathrm{P}(\{X(s+t) = j\} \textstyle\bigcap A) \\
&= \lim_{N \to \infty} \mathrm{P}(\{X^{(N)}(s+t) = j\} \textstyle\bigcap A \bigcap \{\zeta_N > s\}) \\
&= \lim_{N \to \infty} (\boldsymbol{P}^{(N)}(t))_{ij} \mathrm{P}(A \textstyle\bigcap \{\zeta_N > s\}) = (\boldsymbol{P}(t))_{ij} \mathrm{P}(A).
\end{aligned}$$

当然, 一旦我们知道 (4.2.4) 式成立, 则和在有界情形得出 (4.1.9) 式完全一样的讨论, 可以证明这里的 $\{\boldsymbol{P}(t):t\geqslant 0\}$ 也是一个半群.

为了验证 (4.3.4) 式, 把 (4.2.7) 应用于 $\{\boldsymbol{P}^{(N)}(t):t\geqslant 0\}$. 当 N 足够大使得 $i\in F_N$ 时, 有

$$(\boldsymbol{P}^{(N)}(t))_{ij}=\delta_{i,j}+\int_0^t(\boldsymbol{QP}^{(N)}(\tau))_{ij}d\tau.$$

由于 $1\geqslant(\boldsymbol{P}^{(N)}(\tau))_{kj}\to(\boldsymbol{P}(\tau))_{kj}$, 而 $\sum\limits_k|(\boldsymbol{Q})_{ik}|=2R_i<\infty$, 因此由 Lebesgue 控制收敛定理可得 (4.3.4) 成立. □

4.3.2. 非爆炸或爆炸的准则: 本小节我们将首先推导两个保证非爆炸, 即 $e=\infty$ 以概率 1 成立的准则. 我们也将给出一个确保以概率 1 爆炸的条件.

4.3.5 定理. *设 $\boldsymbol{P}$ 是一个转移概率矩阵, 满足 $(\boldsymbol{P})_{ii}=0$, $i\in\mathbb{S}$. 设 $\boldsymbol{\mu}$ 是一个概率向量, 除非 i 关于 $\boldsymbol{P}$ 是常返的, 否则 $(\boldsymbol{\mu})_i=0$. 那么对任一选定的速率 $\mathcal{R}$, 存在一个由 (4.2.1) 式给定的具有初始分布 $\boldsymbol{\mu}$ 的非爆炸过程.*

证明: 首先指出只需处理非退化速率的情形就够了. 事实上, 要验证的是 $J_\infty=\sum\limits_1^\infty R_{X_m}^{-1}E_m=\infty$ 以概率 1 成立. 很明显只有使爆炸的可能性愈小才能使速率愈小. 此外, 因为 $\mathrm{P}(e=\infty)=\sum\limits_{i\in\mathbb{S}}(\boldsymbol{\mu})_i\mathrm{P}(e=\infty|X(0)=i)$, 我们只要证明当 i 关于 $\boldsymbol{P}$ 是常返时, $\mathrm{P}(e=\infty|X(0)=i)=1$ 就够了.

因为我们假定速率是非退化的, (4.2.3) 式保证了被 $\{X(t):t\in[0,J_\infty)\}$ 访问的点与被 $\{X(n):n\geqslant 0\}$ 访问的点一样. 特别地, 对任意 $N\geqslant 1$,

$$\begin{aligned}&\mathrm{P}((\sigma^{(N)})_i<\zeta_N|X(0)=i)\\&=\theta_N\equiv\mathrm{P}(\exists\, n\geqslant 1,\ X_n=i,X_m\in F_N\ ,\ 0\leqslant m\leqslant n|X_0=i),\end{aligned}$$

这里 $(\sigma^{(N)})_i$ 是 $J_1^{(N)}$ 后使得 $t\rightsquigarrow X^{(N)}(t)$ 首次回到 i 的时刻. 与此同时, 基本上与证明定理 2.3.6 的论证一样可证明 $\mathrm{P}((\sigma^{(N)})_i^{(m)}<\zeta_N|X(0)=i)=\theta_N^m$, 其中 $(\sigma^{(N)})_i^{(m)}$ 按归纳定义, 使得 $(\sigma^{(N)})_i^{(1)}=(\sigma^{(N)})_i$,

$$(\sigma^{(N)})_i^{(m+1)}=\begin{cases}\inf\{t\geqslant J_{l+1}^{(N)}:X^{(N)}=i\}, & \text{若 } (\sigma^{(N)})_i^{(m)}=J_l^{(N)}<\infty;\\ \infty, & \text{若 } (\sigma^{(N)})_i^{(m)}=\infty\end{cases}$$

且 $\{J_n^{(N)}:n\geqslant 0\}$ 是 $t\rightsquigarrow X^{(N)}(t)$ 的跳跃时刻. 因此, 如果 i 关于 $\boldsymbol{P}$ 是

常返的, 并由此当 $N\to\infty$ 时 $\theta_N\to 1$, 那么对每一 $m\geqslant 1$,

$$\lim_{N\to\infty}\mathrm{P}((\sigma^{(N)})_i^{(m)}<\zeta_N|X(0)=i)=1.$$

另一方面, 因为 $(\sigma^{(N)})_i^{(1)}\geqslant J_1^{(N)}$ 和

$$(\sigma^{(N)})_i^{(m)}=J_l^{(N)}\implies(\sigma^{(N)})_i^{(m+1)}-(\sigma^{(N)})_i^{(m)}\geqslant J_{l+1}^{(N)}-J_l^{(N)}=\frac{E_{l+1}}{R_i},$$

所以

$$\mathrm{P}((\sigma^{(N)})_i^{(m)}\leqslant T|X(0)=i)\leqslant\mathrm{P}\left(\sum_{l=1}^m E_l\leqslant R_iT\right)\leqslant\frac{(TR_i)^m}{m!}.$$

但是, 对所有 $m\geqslant 1$ 和 $N\geqslant 1$,

$$\begin{aligned}\mathrm{P}(\zeta_N\leqslant T|X(0)=i)\leqslant\mathrm{P}&\left(\zeta_N\leqslant(\sigma^{(N)})_i^{(m)}|X(0)=i\right)+\\&\mathrm{P}\left((\sigma^{(N)})_i^{(m)}\leqslant T|X(0)=i\right),\end{aligned}$$

因此, 先让 $N\to\infty$, 再令 $m\to\infty$, 我们看到当 i 关于 $\boldsymbol{P}$ 是常返时, $\mathrm{P}(\boldsymbol{e}\leqslant T|X(0)=i)=0$. □

另一证明: 同样只需处理非退化速率的情形, 并且只要证明当 i 关于 $\boldsymbol{P}$ 是常返时, $\mathrm{P}(\boldsymbol{e}=\infty|X(0)=i)=1$. 速率的非退化性和 (4.2.3) 式保证了被 $\{X(t):t\in[0,J_\infty)\}$ 访问的点与被 $\{X(n):n\geqslant 0\}$ 访问的点一样. 特别地, 对任意 $T>0$ 和 $m\geqslant 1$ 有

$$\begin{aligned}\mathrm{P}(J_n\leqslant T|X(0)=i)\leqslant\mathrm{P}(J_n\leqslant T,\ &\rho_i^{(m)}\leqslant n|X(0)=i)+\\&\mathrm{P}(\rho_i^{(m)}>n|X(0)=i),\end{aligned}$$

其中 $\rho_i^{(m)}$ 是 $\{X_n:n\geqslant 0\}$ 第 m 次返回 i 的时刻. 因为 i 是 $\{X_n:n\geqslant 0\}$ 的常返态, 所以当 $n\to\infty$ 时上式第二项趋于 0. 同时

$$\rho_i^{(m)}\leqslant n\implies J_n\geqslant\frac{1}{R_i}\sum_{l=1}^m E_{\rho_i^{(l)}}.$$

从而第一项被 m 个相互独立的单位指数随机变量之和小于或等于 R_iT 这一事件的概率所控制, 并且这一概率当 $m\to\infty$ 时趋于 0. 因此我们得证: 对所有的 $T>0$,

$$\mathrm{P}(J_\infty\leqslant T|X(0)=i)=\lim_{n\to\infty}\mathrm{P}(J_n\leqslant T|X(0)=i)=0.$$

于是

$$\mathrm{P}(e<\infty|X(0)=i)=\mathrm{P}(J_\infty<\infty|X(0)=i)$$
$$=\lim_{T\to\infty}\mathrm{P}(J_\infty\leqslant T|X(0)=i)=0.\quad\square$$

我们把第二个非爆炸准则和爆炸准则组合在一起构成同一定理的两个部分. 在这个定理中, 由 (4.2.1) 决定的过程具有速率 $\mathcal{R}$ 和转移概率 $\boldsymbol{P}$.

4.3.6 定理. 如果在 $\mathbb{S}$ 上存在一非负函数 u 使得当 $N\to\infty$ 时, $U_N\equiv\inf\limits_{j\notin F_N}u(j)\to\infty$, 且对某一 $\alpha\in[0,\infty)$, 只要 $i\in\mathbb{S}$ 和 $R_i>0$,

$$\sum_{j\in\mathbb{S}}(\boldsymbol{P})_{ij}u(j)\leqslant\left(1+\frac{\alpha}{R_i}\right)u(i),$$

那么对所有 $i\in\mathbb{S}$, $\mathrm{P}(e=\infty|X(0)=i)=1$. 另一方面, 如果对某一 $i\in\mathbb{S}$, 只要 $i\to j$, 就有 $R_j>0$, 且在 $\mathbb{S}$ 上存在一非负函数 u, 具有性质: 对某一 $\epsilon>0$, 只要 $i\to j$,

$$\sum_{\{k:i\to k\}}(\boldsymbol{P})_{jk}u(k)\leqslant u(j)-\frac{\epsilon}{R_j},$$

那么 $\mathrm{P}(e=\infty|X(0)=i)=0$.

证明: 为了证明第一部分, 对每一 $N\geqslant1$, 当 $j\in F_N$ 时, 令 $u^{(N)}(j)=u(j)$, 当 $j\notin F_N$ 时, 令 $u^{(N)}(j)=U_N$. 容易证明: 如果 $\boldsymbol{Q}^{(N)}=\boldsymbol{R}^{(N)}(\boldsymbol{P}-\boldsymbol{I})$, 其中 $(\boldsymbol{R}^{(N)})_{ij}=R_i^{(N)}\delta_{i,j}$, 且 $\boldsymbol{u}^{(N)}$ 是由 $(\boldsymbol{u}^{(N)})_i=u^{(N)}(i)$ 决定的列向量, 则对所有 $i\in\mathbb{S}$, $(\boldsymbol{Q}^{(N)}\boldsymbol{u}^{(N)})_i\leqslant\alpha u^{(N)}(i)$. 因此把 Kolmogorov 向前方程 (4.2.9) 应用到 $\{\boldsymbol{P}^{(N)}(t):t\geqslant0\}$ 上, 就有

$$\frac{d}{dt}\Big(\boldsymbol{P}^{(N)}(t)\boldsymbol{u}^{(N)}\Big)_i\leqslant\alpha\Big(\boldsymbol{P}^{(N)}(t)\boldsymbol{u}^{(N)}\Big)_i,$$

从而, $(\boldsymbol{P}^{(N)}(T)\boldsymbol{u}^{(N)})_i\leqslant e^{\alpha T}u^{(N)}(i)$. 但如果 $\zeta_N\leqslant T$, 则

$$u^{(N)}(X^{(N)}(T))=u^{(N)}(X^{(N)}(\zeta_N))\geqslant U_N,$$

这意味着, 对 $i\in F_N$,

$$\boldsymbol{P}(\zeta_N\leqslant T|X(0)=i)\leqslant\frac{1}{U_N}\mathrm{E}\left[u^{(N)}(X^{(N)}(T))|X(0)=i\right]$$
$$=\frac{(\boldsymbol{P}^{(N)}(T)\boldsymbol{u}^{(N)})_i}{U_N}\leqslant\frac{e^{\alpha T}u^{(N)}(i)}{U_N}\leqslant\frac{e^{\alpha T}u(i)}{U_N}.$$

因此由单调收敛定理, 我们已经证明了 $\mathrm{P}(e \leqslant T|X(0)=i) \leqslant \lim_{N\to\infty}\mathrm{P}(\zeta_N \leqslant T|X(0)=i)=0$.

在证明第二个结论的过程中, 我们可以假设对所有 $j\in\mathbb{S}$ 有 $i\to j$. 取 $u^{(N)}(j)=u(j)$, $j\in F_N$; $u^{(N)}(j)=0$, $j\notin F_N$. 用 $\boldsymbol{u}^{(N)}$ 表示相应的列向量. 很明显, 根据 $i\in F_N$ 或 $i\notin F_N$, $(\boldsymbol{Q}^{(N)}\boldsymbol{u}^{(N)})_i$ 要么小于等于 $-\epsilon$ 要么等于 0. 因此再次利用 Kolmogorov 向前方程, 得

$$\frac{d}{dt}\left(\boldsymbol{P}^{(N)}(t)\boldsymbol{u}^{(N)}\right)_i \leqslant -\epsilon\sum_{j\in F_N}\left(\boldsymbol{P}^{(N)}(t)\right)_{ij}.$$

从而

$$\begin{aligned}
&\mathrm{E}\left[u^{(N)}(X^{(N)}(T))|X(0)=i\right]-u^{(N)}(i)\\
\leqslant\ & -\epsilon\mathrm{E}\left[\int_0^T \mathbf{1}_{F_N}(X^{(N)}(t))dt|X^{(N)}(0)=i\right]\\
=\ & -\epsilon\mathrm{E}[T\wedge\zeta_N|X^{(N)}(0)=i].
\end{aligned}$$

但是因为 $u^{(N)}\geqslant 0$, 这意味着对所有 N, $\mathrm{E}[\zeta_N|X^{(N)}(0)=i]\leqslant\dfrac{u(i)}{\epsilon}$, 所以 $\mathrm{E}[e|X(0)=i]\leqslant\dfrac{u(i)}{\epsilon}<\infty$. □

4.3.3. 当爆炸发生时做什么: 虽然我无意后悔, 但不得不承认我们已经忽略了从数学观点来看所考虑的理论最有兴趣的方面. 也就是说, 直到现在还没有阐述当爆炸以正概率发生时应有的对策. 在本节我们将仅讨论众多可供选择的最平凡的一种.

如果把 $\{e<\infty\}$ 作为在有限时间内过程逃出其状态空间的事件, 则在 e 之后过程的延续可能有必要引入至少一个*新状态*. 事实上, 这里的情况和拓扑学家在他们想要紧化一个空间时遇到的情形非常相似. 一个可分局部紧空间的最简单紧化是*单点紧化*. 这种紧化承认逃向无穷但忽略所走路径的所有细节. 在目前情况下单点紧化的类比是, 在爆炸的时刻, 把过程送到一个吸收点 $\Delta\notin\mathbb{S}$. 也就是说, 定义 $t\in[0,\infty)\mapsto X(t)\in\mathbb{S}\bigcup\{\Delta\}$ 如下: 只要 $t\in[0,J_\infty)$, 即可根据在 (4.2.3) (记住, 由定理 4.3.1, 以概率 1 有 $e=J_\infty$) 中的规定定义, 而当 $t\in[J_\infty,\infty)$ 时, 取 $X(t)=\Delta$. 由于各种原因, 我将不做任何解释地称这种扩张为过程的*最小扩张*. 这种最小扩张具有总能进行的优点. 但另一方面, 像单点紧化一样, 它的缺点是完全掩盖了所有任何特殊情形的精细结构. 例如, 当 $\mathbb{S}=\mathbb{Z}$ 时, 爆炸可能发生是因为给定 $e<\infty$, 以概率 1 成立

$\lim_{t\nearrow e} X(t) = +\infty$. 也可以是因为, 虽然以概率 1 成立 $\lim_{t\nearrow e} |X(t)| = +\infty$, 但是 $\lim_{t\nearrow e} X(t) = +\infty$ 和 $\lim_{t\nearrow e} X(t) = -\infty$ 都以正概率发生. 在第二种情形, 人们可能想记录这两种可能性发生的大小, 而这可通过引入两个新的吸收状态 Δ_+ 和 Δ_- 来完成, Δ_+ 对应经 $+\infty$ 爆发的那些路径, Δ_- 对应经 $-\infty$ 爆发的那些路径.

换句话说, 不是把爆炸时间看做过程从 $\mathbb{S}$ 溢出的时间, 而是把它看成在时刻 e 重新在 $\mathbb{S}$ 上区分过程直到它再一次爆炸的更新时间, 等等.

显然有无穷多种可能性. 这足以说明前面的讨论几乎只抓住了问题的表面.

4.4 遍历性质

本节我们将检查本章已讨论过的 Markov 过程的遍历行为. 所考虑的是具有在 §4.2.1 中描述的那种连续时间 Markov 过程, 假设的条件是没有爆炸.

4.4.1. 状态的分类: 正如在 §3.1 一样, 我们从对 $\mathbb{S}$ 的状态分类开始.

为了阐述我们的第一个观点, 记 $\boldsymbol{Q} = \boldsymbol{R}(\boldsymbol{P} - \boldsymbol{I})$, 这里 $\boldsymbol{R}$ 是来自 $\mathcal{R}$ 的速率的对角线矩阵, $\boldsymbol{P}$ 是对角线元素全为 0 的转移概率矩阵. 下面定义 $\boldsymbol{P}^{\mathcal{R}}$ 为由[5]

$$(\boldsymbol{P}^{\mathcal{R}})_{ij} = \begin{cases} (\boldsymbol{P})_{ij}, & \text{若 } R_i > 0; \\ \delta_{ij}, & \text{若 } R_i = 0 \end{cases} \tag{4.4.1}$$

给定的转移概率矩阵. 显然 $\boldsymbol{Q} = \boldsymbol{R}(\boldsymbol{P}^{\mathcal{R}} - \boldsymbol{I})$. 此外,

$$\begin{aligned} \text{关于 } \boldsymbol{P}^{\mathcal{R}}, i \to j &\iff \exists n \geqslant 0,\ (\boldsymbol{Q}^n)_{ij} > 0 \\ &\iff \text{对所有 } t > 0,\ \ (\boldsymbol{P}(t))_{ij} > 0. \end{aligned} \tag{4.4.2}$$

根据 (4.2.13) 式, 当 $\mathcal{R}$ 有界时, 这是显然的; 一般情形可从有界情形和 (参看 §4.3.1 中的记号) $\lim_{N\to\infty} (\boldsymbol{P}^{(N)}(t))_{ij} = (\boldsymbol{P}(t))_{ij}$ 得到.

基于 (4.4.2) 式, 当其中任意一个等价条件成立时, 我们记作 $i \xrightarrow{\boldsymbol{Q}} j$. 当 $i \xrightarrow{\boldsymbol{Q}} j$ 且 $j \xrightarrow{\boldsymbol{Q}} i$ 时, 称 i 与 j 是 ***Q*-互通** 的, 并记作 $i \overset{\boldsymbol{Q}}{\leftrightarrow} j$. 与此相联系, 如果任一状态与所有其他的状态互通, 则称 $\mathbb{S}$ 是 ***Q*-不可约的**.

[5] 值得注意的是, 不同于 $\boldsymbol{P}$, $\boldsymbol{P}^{\mathcal{R}}$ 是完全由 $\boldsymbol{Q}$ 决定的.

下面我们转向常返性和瞬时性问题. 因为

$$R_i = 0 \implies \mathrm{P}\,(\text{对所有}\ \ t \geqslant 0, X(t) = i|X(0) = i) = 1,$$

所以很显然, 当 $R_i = 0$ 时, i 应被认为是常返的. 另一方面, 由于在连续时间背景下不存在 "第一次正的时刻", 当 $R_i > 0$ 时, i 的常返性的含义是不明显的. 然而, 如果采取这样的看法: 在离散时间情形下 1 时刻表示链能够移动的第一时间, 则很清楚 1 应该起第一次跳跃时刻 J_1 的作用, 从而 ρ_i 的作用由

$$\sigma_i = \inf\{t \geqslant J_1 : X(t) = i\} \tag{4.4.3}$$

来扮演. 因此, 如果 $R_i = 0$ 或 $\mathrm{P}(\sigma_i < \infty|X(0) = i) = 1$, 则我们说 i 是 **$\boldsymbol{Q}$-常返的**; 如果 i 不是 $\boldsymbol{Q}$-常返的, 则称其是 **$\boldsymbol{Q}$-瞬时的**.

下面我们考察一个事实: *i 是 $\boldsymbol{Q}$-常返的当且仅当它关于 (4.4.1) 中的转移概率矩阵 $\boldsymbol{P}^{\mathcal{R}}$ 是常返的*. 实际上, 从 §4.2.1 的讨论易知, 被 $t \rightsquigarrow X(t)$ 访问的点恰恰与被 $n \rightsquigarrow X_n = X(J_n)$ 访问的点是完全一样的, 而 $\{X_n : n \geqslant 0\}$ 是具有转移概率矩阵 $\boldsymbol{P}^{\mathcal{R}}$ 的 Markov 链. 作为上述事实的一个推论, 我们看到不仅定理 3.1.2 而且推论 3.1.4 都可运用到现在的背景中. 特别地, *$\boldsymbol{Q}$-常返性和 $\boldsymbol{Q}$-瞬时性是 $\boldsymbol{Q}$-互通类性质*.

我们下面推导一个在连续时间状态下与 (2.3.7) 式类似的关系. 令 $\sigma_j^{(1)} = \sigma_j$, 且对 $m \geqslant 1$, 若 $\sigma_j^{(m)} = \infty$, 则令 $\sigma_j^{(m+1)} = \infty$; 若 $\sigma_j^{(m)} = J_\ell < \infty$, 则令 $\sigma_j^{(m+1)} = \inf\{t \geqslant J_{\ell+1} : X(t) = j\}$. 因此正如定理 2.3.6 的证明一样, 有

$$\mathrm{P}\left(\sigma_i^{(m)} < \infty|X(0) = i\right) = \mathrm{P}\,(\sigma_i < \infty|X(0) = i)^m.$$

另外, 容易验证

$$\begin{aligned}
&\mathrm{E}\left[\int_{\sigma_i^{(m)}}^{\sigma_i^{(m+1)}} \mathbf{1}_{\{i\}}(X(t))dt,\ \sigma_i^{(m)} < \infty\middle|X(0) = i\right] \\
&= \mathrm{E}\left[\int_0^{\sigma_i} \mathbf{1}_{\{i\}}(X(t))dt\middle|X(0) = i\right]\mathrm{P}(\sigma_i^{(m)} < \infty|X(0) = i) \\
&= \mathrm{E}[J_1|X(0) = i]\mathrm{P}(\sigma_i^{(m)} < \infty|X(0) = i) = \frac{\mathrm{P}(\sigma_i < \infty|X(0) = i)^m}{R_i}
\end{aligned}$$

和

$$\begin{aligned}&\mathrm{E}\left[\int_0^\infty \mathbf{1}_{\{j\}}(X(t))dt \Big| X(0)=i\right]\\ =&\mathrm{E}\left[\int_0^\infty \mathbf{1}_{\{j\}}(X(t))dt \Big| X(0)=j\right]\mathrm{P}(\sigma_j<\infty|X(0)=i).\end{aligned}$$

因此, 利用从定理 2.3.6 推导 (2.3.7) 式完全一样的论证, 可得

$$\begin{aligned}&\mathrm{E}\left[\int_0^\infty \mathbf{1}_{\{j\}}(X(t))dt \Big| X(0)=i\right]\\ &=\frac{1}{R_j}\left(\delta_{i,j}+\frac{\mathrm{P}(\sigma_j<\infty|X(0)=i)}{\mathrm{P}(\sigma_j=\infty|X(0)=j)}\right),\\ &\mathrm{E}\left[\int_0^\infty \mathbf{1}_{\{i\}}(X(t))dt \Big| X(0)=i\right]=\infty\\ \Longleftrightarrow\ &\mathrm{P}\left(\int_0^\infty \mathbf{1}_{\{i\}}(X(t))dt=\infty \Big| X(0)=i\right)=1,\\ &\mathrm{E}\left[\int_0^\infty \mathbf{1}_{\{i\}}(X(t))dt \Big| X(0)=i\right]<\infty\\ \Longleftrightarrow\ &\mathrm{P}\left(\int_0^\infty \mathbf{1}_{\{i\}}(X(t))dt<\infty \Big| X(0)=i\right)=1.\end{aligned} \tag{4.4.4}$$

本节的最后一个目标是证明下列结论.

4.4.5 定理. *对任意给定的状态 $i\in\mathbb{S}$, 下列陈述等价:*

(1) *i 是 $\boldsymbol{Q}$-常返的.*

(2) *存在一个 $t\in(0,\infty)$ 使得 i 关于转移概率矩阵 $\boldsymbol{P}(t)$ 是常返的.*

(3) *对所有 $t\in(0,\infty)$, i 关于转移概率矩阵 $\boldsymbol{P}(t)$ 是常返的.*

证明: 我们将通过验证当 “常返” 全部被 “瞬时” 代替后相同的表述成立来证明这种等价性. 为此, 首先指出

$$\mathrm{E}\left[\int_0^\infty \mathbf{1}_{\{i\}}(X(t))dt \Big| X(0)=i\right]=\int_0^\infty (\boldsymbol{P}(t))_{ii}dt,$$

因此从 (4.4.4) 式的第一行得到

$$i \text{ 是 } \boldsymbol{Q}\text{-瞬时的} \Longleftrightarrow \int_0^\infty (\boldsymbol{P}(t))_{ii}dt<\infty. \tag{4.4.6}$$

其次, 因为 $(\boldsymbol{P}(h))_{ii}\geqslant \mathrm{P}(J_1>h|X(0)=i)=e^{-hR_i}$, 故对 $0\leqslant s<t$,

$$(\boldsymbol{P}(t))_{ii}\geqslant(\boldsymbol{P}(t-s))_{ii}(\boldsymbol{P}(s))_{ii}\geqslant e^{-(t-s)R_i}(\boldsymbol{P}(s))_{ii}. \tag{4.4.7}$$

因此对任意 $t>0$ 和 $n\in\mathbb{N}$,

$$(\boldsymbol{P}(t)^{n+1})_{ii}=(\boldsymbol{P}((n+1)t))_{ii}\geqslant e^{-tR_i}(\boldsymbol{P}(\tau))_{ii}$$

对所有的 $\tau\in[nt,(n+1)t]$ 成立, 并且

$$e^{-tR_i}(\boldsymbol{P}(t)^n)_{ii}=e^{-tR_i}(\boldsymbol{P}(nt))_{ii}\leqslant(\boldsymbol{P}(\tau))_{ii}$$

对所有的 $\tau\in[nt,(n+1)t]$ 成立. 由此可得

$$te^{-tR_i}\sum_{n=0}^{\infty}(\boldsymbol{P}(t)^n)_{ii}\leqslant\int_0^{\infty}(\boldsymbol{P}(\tau))_{ii}d\tau\leqslant te^{tR_i}\sum_{n=0}^{\infty}(\boldsymbol{P}(t)^{n+1})_{ii},$$

于是欲证的等价关系立即从 (4.4.6) 式得到. □

4.4.2. 平稳测度与极限定理: 在这一小节, 我们将证明下列基本结果.

4.4.8 定理. 对每一 $j\in\mathbb{S}$,

$$\hat{\pi}_{jj}\equiv\lim_{t\to\infty}(\boldsymbol{P}(t))_{jj}\ \text{存在},$$

且对 $i\neq j$,

$$\lim_{t\to\infty}(\boldsymbol{P}(t))_{ij}=\hat{\pi}_{ij}\equiv\mathrm{P}(\sigma_j<\infty|X(0)=i)\hat{\pi}_{jj}.$$

进而, 如果 $\hat{\pi}_{jj}>0$, 则对所有 $i\in C\equiv\{i:i\overset{\boldsymbol{Q}}{\leftrightarrow}j\}$, $\hat{\pi}_{ii}>0$; 当行向量 $\hat{\boldsymbol{\pi}}^C$ 由 $(\hat{\boldsymbol{\pi}}^C)_i=\mathbf{1}_C(i)\hat{\pi}_{ii}$ 决定时, 那么, 对每一 $s>0$, $\hat{\boldsymbol{\pi}}^C$ 是唯一的概率向量 $\boldsymbol{\mu}\in\mathrm{Stat}(\boldsymbol{P}(s))$, 使得对 $k\notin C$, $(\boldsymbol{\mu})_k=0$. 事实上, 如果对某一 $s>0$, $\boldsymbol{\mu}\in\mathrm{Stat}(\boldsymbol{P}(s))$, 则对每一 $j\in C$, 有

$$(\boldsymbol{\mu})_j=\left(\sum_{i\overset{\boldsymbol{Q}}{\leftrightarrow}j}(\boldsymbol{\mu})_i\right)\hat{\pi}_{jj}.$$

证明: 我们从下列连续时间的更新方程开始 (参看 (3.2.6)):

$$(\boldsymbol{P}(t))_{ij}=e^{-tR_i}\delta_{i,j}+\mathrm{E}[(\boldsymbol{P}(t-\sigma_j))_{jj},\sigma_j\leqslant t|X(0)=i].\qquad(4.4.9)$$

(4.4.9) 式的证明如下. 首先, 把 $(\boldsymbol{P}(t))_{ij}$ 写成

$$\mathrm{P}(X(t)=j,\ J_1>t|X(0)=i)+\mathrm{P}(X(t)=j,J_1\leqslant t|X(0)=i).$$

很明显, 除非 $i=j$, 右边第一项为 0, 而 $i=j$ 时, 它等于 e^{-tR_i}. 为处理第二项, 把它写成

$$\sum_{m=1}^{\infty} \mathrm{P}(X(t)=j,\sigma_j=J_m\leqslant t|X(0)=i),$$

注意到 (参看 (4.2.3) 和 (4.2.5))

$$\begin{aligned}
&\mathrm{P}(X(t)=j,\sigma_j=J_m\leqslant t|X(0)=i)\\
=&\mathrm{P}(\Phi^{\mathcal{R}}(t-J_m;(E_{m+1},\cdots,E_{m+n},\cdots),(j,X_{m+1},\cdots,\\
&X_{m+n},\cdots))=j,\sigma_j=J_m\leqslant t|X(0)=i)\\
=&\boldsymbol{E}\left[(\boldsymbol{P}(t-J_m))_{jj},\sigma_j=J_m\leqslant t|X(0)=i\right].
\end{aligned}$$

因此, 将上式关于 $m\geqslant 1$ 求和, 并与前面的结果相结合即可得到 (4.4.9) 式.

证明了 (4.4.9) 式之后, 为了证明定理的第一部分, 我们只需处理 $i=j$ 的情形即可. 为此, 首先注意到, 由 (4.4.7) 式, 对所有 $s>0$ 和 $i\in\mathbb{S}$, $(\boldsymbol{P}(s))_{ii}\geqslant e^{-sR_i}>0$, 所以每个 i 关于 $\boldsymbol{P}(s)$ 都是非周期的. 因此由 (3.2.15) 式, 对所有 $s>0$ 和 $i\in\mathbb{S}$, $\pi(s)_{ii}\equiv\lim\limits_{n\to\infty}(\boldsymbol{P}(s)^n)_{ii}$ 存在. 我们想证明 $\pi(1)_{ii}=\lim\limits_{t\to\infty}(\boldsymbol{P}(t))_{ii}$. 而当 $\pi(1)_{ii}=0$ 时, 易证结论成立. 事实上, 由 (4.4.7) 式,

$$\limsup_{t\to\infty}(\boldsymbol{P}(t))_{ii}\leqslant e^{R_i}\limsup_{t\to\infty}(\boldsymbol{P}([t]+1))_{ii}=e^{R_i}\pi(1)_{ii},$$

其中 $[t]$ 表示 t 的整数部分.

根据前面的证明, 对于第一部分结论只剩下证明: 当 $\pi(1)_{ii}>0$ 时, 有 $\lim\limits_{t\to\infty}(\boldsymbol{P}(t))_{ii}=\pi(1)_{ii}$. 而关键一步将是证明对所有 $s>0$, $\pi(s)_{ii}=\pi(1)_{ii}$. 我们已知当 $\pi(1)_{ii}=0$ 时结论成立. 因此假定 $\pi(1)_{ii}>0$. 设 C 是 i 的关于 $\boldsymbol{P}(1)$ 的 $\boldsymbol{Q}$-互通类. 由 (4.4.2) 式, 对每一 $s>0$, C 也是 i 的关于 $\boldsymbol{P}(s)$ 的互通类. 此外, 由于 $\pi(1)_{ii}>0$, i 关于 $\boldsymbol{P}(1)$ 是常返的, 事实上是正常返的, 因此, 由定理 4.4.5, 它对所有 $s>0$ 相对于 $\boldsymbol{P}(s)$ 也是常返的. 现在定义行向量 $\boldsymbol{\pi}(1)^C$ 使得 $(\boldsymbol{\pi}(1)^C)_j=\mathbf{1}_C(j)\pi(1)_{jj}$. 那么因为 $\pi(1)_{ii}>0$, 由定理 3.2.10, 我们知道 $\boldsymbol{\pi}(1)^C$ 是唯一的在 C 之外为 0 的 $\boldsymbol{\mu}\in\mathrm{Stat}(\boldsymbol{P}(1))$. 下面给定 $s>0$, 考虑 $\boldsymbol{\mu}=\boldsymbol{\pi}(1)^C\boldsymbol{P}(s)$. 从而 $\boldsymbol{\mu}$ 是一个概率向量, 并且由于当 $j\in C$ 和 $k\notin C$ 时, 有 $(\boldsymbol{P}(1))_{jk}=0$,

所以 μ 在 C 之外为 0. 又因为

$$\begin{aligned}\boldsymbol{\mu}\boldsymbol{P}(1) &= \boldsymbol{\pi}(1)^C\boldsymbol{P}(s)\boldsymbol{P}(1) = \boldsymbol{\pi}(1)^C\boldsymbol{P}(s+1)\\ &= \boldsymbol{\pi}(1)^C\boldsymbol{P}(1)\boldsymbol{P}(s) = \boldsymbol{\pi}(1)^C\boldsymbol{P}(s) = \boldsymbol{\mu},\end{aligned}$$

所以 $\boldsymbol{\mu}$ 关于 $\boldsymbol{P}(1)$ 是平稳的. 因此由唯一性, 我们得到 $\boldsymbol{\pi}(1)^C\boldsymbol{P}(s) = \boldsymbol{\pi}(1)^C$, 从而 $\boldsymbol{\pi}(1)^C$ 是一个在 C 之外为 0的关于 $\boldsymbol{P}(s)$ 的平稳测度. 但由 (3.2.8) 式并注意到 C 也是 i 关于 $\boldsymbol{P}(s)$ 的互通类, 我们有

$$\pi(1)_{ii} = (\boldsymbol{\pi}(1)^C)_i = \left(\sum_{j\in C}(\boldsymbol{\pi}(1)^C)_j\right)\pi(s)_{ii} = \pi(s)_{ii}.$$

为了根据上式完成第一部分的证明, 注意到, 由 (4.4.7) 式, 当 $ns \leqslant t \leqslant (n+1)s$ 时, 有

$$e^{-sR_j}(\boldsymbol{P}(ns))_{jj} \leqslant (\boldsymbol{P}(t))_{jj} \leqslant e^{sR_j}(\boldsymbol{P}((n+1)s))_{jj} \quad .$$

又因为 $\pi(s)_{jj} = \pi(1)_{jj}$ 且 $\boldsymbol{P}(nt) = \boldsymbol{P}(t)^n$, 所以

$$e^{-sR_j}\pi(1)_{jj} \leqslant \liminf_{t\to\infty}(\boldsymbol{P}(t))_{jj} \leqslant \limsup_{t\to\infty}(\boldsymbol{P}(t))_{jj} \leqslant e^{sR_j}\pi(1)_{jj}.$$

令 $s \searrow 0$, 并定义 $\hat{\pi}_{jj} = \pi(1)_{jj}$.

给出第一部分的证明后, 第二部分的证明就容易了. 如果 $\hat{\pi}_{jj} > 0$, $C = \{i : i \overset{Q}{\leftrightarrow} j\}$, 则对每一 $s > 0$, 我们知道 C 是 j 的关于 $\boldsymbol{P}(s)$ 的互通类且 $\hat{\boldsymbol{\pi}}^C = \boldsymbol{\pi}(s)^C \in \mathrm{Stat}(\boldsymbol{P}(s))$. 反之, 由 (3.2.8) 式, 我们得到, 如果对某一 $s > 0$, $\boldsymbol{\mu} \in \mathrm{Stat}(\boldsymbol{P}(s))$, 则

$$(\boldsymbol{\mu})_j = \left(\sum_{\{i:i\overset{Q}{\leftrightarrow}j\}}(\boldsymbol{\mu})_i\right)\pi(s)_{jj} = \left(\sum_{\{i:i\overset{Q}{\leftrightarrow}j\}}(\boldsymbol{\mu})_i\right)\hat{\pi}_{jj}. \quad \square$$

根据前面的结果, 如果 $\hat{\pi}_{ii} > 0$, 我们称 i 是 *Q-正常返的*; 如果 i 是 Q-常返的, 但不是 Q-正常返的, 我们称 i 是 *Q-零常返的*. 从前面讨论, 我们已经知道, *Q-正常返性是一种 Q-互通类性质*.

再稍加推导容易得到下列推论.

4.4.10 推论. 平均遍历定理 *假定 j 是 Q-正常返的, 且* $\mathrm{P}(X(0) \overset{Q}{\leftrightarrow} j) = 1$, *则*

$$\lim_{T\to\infty}\mathrm{E}\left[\left(\frac{1}{T}\int_0^T \mathbf{1}_{\{j\}}(X(t))dt - \hat{\pi}_{jj}\right)^2\right] = 0.$$

更精确的结论请看习题 4.5.10.

证明: 证明实际上只是定理 3.2.14 的论证在连续背景下的一种明显的翻版. 记 $C = \{i : j \overset{Q}{\leftrightarrow} i\}$. 正如那里已经给出的论证一样, 只需处理当 $\hat{\boldsymbol{\pi}}^C$ 是初始分布的情形. 因此, 如果 $f = \mathbf{1}_{\{j\}} - \hat{\pi}_{jj}$, 需要证明当 $\hat{\boldsymbol{\pi}}^C$ 是 $X(0)$ 的分布时, 有

$$\lim_{T\to\infty} \mathrm{E}\left[\left(\frac{1}{T}\int_0^T f(X(t))dt\right)^2\right] = 0.$$

但由于 $\hat{\boldsymbol{\pi}}^C$ 是 $\boldsymbol{P}(t)$-平稳的, 所以有

$$\begin{aligned}\mathrm{E}\left[\left(\frac{1}{T}\int_0^T f(X(t))dt\right)^2\right] &= \frac{2}{T^2}\int_0^T \left(\int_0^t \mathrm{E}\left[f(X(s))f(X(t))\right]ds\right)dt \\ &= \frac{2}{T^2}\int_0^T \left(\int_0^t \alpha(t-s)ds\right)dt \\ &= \frac{2}{T}\int_0^T \left(1-\frac{t}{T}\right)\alpha(t)dt,\end{aligned}$$

这里 $\alpha(t) \equiv \hat{\boldsymbol{\pi}}^C(f\boldsymbol{P}(t)\boldsymbol{f})$, $\boldsymbol{f}$ 是相应于函数 f 的列向量, $f\boldsymbol{P}(t)\boldsymbol{f}$ 是由 $(f\boldsymbol{P}(t)\boldsymbol{f})_i = f(i)(\boldsymbol{P}(t)\boldsymbol{f})_i$ 决定的列向量. 最后, 因为对每一 $i \in C$, $\lim\limits_{t\to\infty}(\boldsymbol{P}(t)\boldsymbol{f})_i = 0$, 所以 $\lim\limits_{t\to\infty}\alpha(t) = 0$. 因此当 $T\to\infty$ 时,

$$\frac{2}{T}\int_0^T \left(1-\frac{t}{T}\right)\alpha(t)dt \leqslant \frac{2}{T}\int_0^T |\alpha(t)|dt \to 0. \quad \square$$

4.4.3. 解释 $\hat{\pi}_{ii}$: 尽管我们已经证明极限 $\hat{\pi}_{ij} = \lim\limits_{t\to\infty}(\boldsymbol{P}(t))_{ij}$ 存在, 但我们还必须给出 $\hat{\pi}_{ii}$ 的一个类似于 (3.2.5) 的表达式. 然而, 正如我们将要看到的那样, 这样的表达式容易从 (4.4.9) 式推出. 对每一 $\alpha > 0$, 令

$$L(\alpha)_{ij} = \alpha\mathrm{E}\left[\int_0^\infty e^{-\alpha t}\mathbf{1}_{\{j\}}(X(t))dt\middle|X(0)=i\right] = \alpha\int_0^\infty e^{-\alpha t}(\boldsymbol{P}(t))_{ij}dt.$$

因为 $\hat{\pi}_{ii} = \lim\limits_{t\to\infty}(\boldsymbol{P}(t))_{ii}$, 由上述第二个等式, 容易验证 $\hat{\pi}_{ii}$ 即为 $\lim\limits_{\alpha\searrow 0}L(\alpha)_{ii}$. 与此同时, 从 (4.4.9) 式得到

$$L(\alpha)_{ii} = \frac{\alpha}{\alpha+R_i} + \mathrm{E}[e^{-\alpha\sigma_i}|X(0)=i]L(\alpha)_{ii}.$$

因此, 我们已经证明

$$\hat{\pi}_{ii} = \begin{cases} 1, & \text{若 } R_i = 0; \\ \dfrac{1}{R_i \mathrm{E}[\sigma_i | X(0) = i]}, & \text{若 } R_i > 0. \end{cases}$$

结合定理 4.4.8 中的第二个等式, 得到

$$\hat{\pi}_{ij} = \begin{cases} \delta_{i,j} + \mathrm{P}(\sigma_j < \infty | X(0) = i), & \text{若 } R_j = 0; \\ \dfrac{\mathrm{P}(\sigma_j < \infty | X(0) = i)}{R_j \mathrm{E}[\sigma_j | X(0) = j]}, & \text{若 } R_j > 0. \end{cases} \tag{4.4.11}$$

当然, 作为 (4.4.11) 式的一个直接推论, 我们知道, i **是正常返的当且仅当要么** $R_i = 0$, **要么** $\mathrm{E}[\sigma_i | X(0) = i] < \infty$.

4.5 习题

4.5.1. 这个习题的目的是给出 (4.1.5) 式的另一种推导方法. 设 $\{E_n : n \geqslant 1\}$ 是一列相互独立的单位指数随机变量序列, 如同 §4.1.1 中那样, 定义 $\{J_n : n \geqslant 1\}$ 和 $\{N(t) : n \geqslant 1\}$. 给定 $0 = t_0 < \cdots < t_\ell$ 和 $0 \leqslant n_1 \leqslant \cdots \leqslant n_\ell$, 用多维积分变量替换公式证明:

$$\begin{aligned} & \mathrm{P}(N(t_1) = n_1, \cdots, N(t_\ell) = n_\ell) \\ = & \mathrm{P}(J_{n_1} \leqslant t_1 \leqslant J_{n_1+1}, \cdots, J_{n_\ell} \leqslant t_\ell \leqslant J_{n_\ell+1}) \\ = & \int \cdots \int_A \exp\left(- \sum_{j=1}^{n_\ell+1} \xi_j \right) d\xi_1 \cdots d\xi_{n_\ell+1} \\ = & \int \cdots \int_B e^{-\eta_{n_\ell+1}} d\eta_1 \cdots d\eta_{n_\ell+1} \\ = & e^{-t_\ell} \prod_{j=1}^{\ell} \mathrm{vol}(\Delta_j) = \prod_{j=1}^{\ell} e^{-(t_j - t_{j-1})} \frac{(t_j - t_{j-1})^{n_j - n_{j-1}}}{(n_j - n_{j-1})!}, \end{aligned}$$

其中

$$A = \left\{ (\xi_1, \cdots, \xi_{n_\ell+1}) \in (0, \infty)^{n_\ell+1} : \sum_{k=1}^{n_j} \xi_k \leqslant t_j < \sum_{k=1}^{n_j+1} \xi_k, 1 \leqslant j \leqslant \ell \right\},$$

$$B = \{ (\eta_1, \cdots, \eta_{n_\ell+1}) \in (0, \infty)^{n_\ell+1} : \eta_i < \eta_{i+1}, 1 \leqslant i < n_\ell, \\ \eta_{n_j} \leqslant t_j \leqslant \eta_{n_j+1}, 1 \leqslant j \leqslant \ell \}$$

$$\Delta_j = \{ (u_1, \cdots, u_{n_j - n_{j-1}}) \in \mathbb{R}^{n_j - n_{j-1}} : t_{j-1} \leqslant u_1 < \cdots < \\ u_{n_j - n_{j-1}} \leqslant t_j \},$$

当 $n_j = 0$ $(1 \leqslant j \leqslant i)$ 时, 作适当的修正.

4.5.2. 设 $\boldsymbol{M}_1$ 和 $\boldsymbol{M}_2$ 是 $M_{u,v}(\mathbb{S})$ 中可交换的元素. 验证对所有 $m \in \mathbb{N}$,

$$(\boldsymbol{M}_1 + \boldsymbol{M}_2)^m = \sum_{l=0}^{m} \binom{m}{\ell} {\boldsymbol{M}_1}^{\ell} {\boldsymbol{M}_2}^{m-\ell},$$

然后证明 (4.2.12) 式.

4.5.3. 给定一个 $\boldsymbol{Q}$-常返状态 i, 设 $C = \{j : i \overset{\boldsymbol{Q}}{\leftrightarrow} j\}$. 证明 $R_i > 0 \implies$ 对所有 $j \in C$, $R_j > 0$.

4.5.4. 在这个习题中, 我们将给出习题 2.4.1 中考虑的问题在连续情形时的一种对应. 为此, 假定 $\mathbb{S}$ 是不可约的, 且关于 $\boldsymbol{Q}$ 是正常返的, $\hat{\boldsymbol{\pi}}$ 表示唯一概率向量, 它关于每一 $\boldsymbol{P}(t)$, $t > 0$, 是平稳的. 其次, 给定一伴随半群 $\{\boldsymbol{P}(t)^{\top} : t \geqslant 0\}$ 使得

$$(\boldsymbol{P}(t)^{\top})_{ij} = \frac{(\hat{\boldsymbol{\pi}})_j (\boldsymbol{P}(t))_{ji}}{(\hat{\boldsymbol{\pi}})_i}.$$

(a) 证明 $\boldsymbol{P}(t)^{\top}$ 是一个转移概率矩阵, 且对每一 $t \geqslant 0$ 有 $\hat{\boldsymbol{\pi}} \boldsymbol{P}(t)^{\top} = \hat{\boldsymbol{\pi}}$. 此外, 验证 $\{\boldsymbol{P}(t)^{\top} : t \geqslant 0\}$ 是由 Q-矩阵 $\boldsymbol{Q}^{\top}$ 决定的半群, 这里

$$(\boldsymbol{Q}^{\top})_{ij} = \frac{(\hat{\boldsymbol{\pi}})_j (\boldsymbol{Q})_{ji}}{(\hat{\boldsymbol{\pi}})_i}.$$

(b) 设 P 和 $\mathrm{P}^{\top}$ 分别表示由相应于 $\boldsymbol{Q}$ 和 $\boldsymbol{Q}^{\top}$ 的具有初始分布 $\hat{\boldsymbol{\pi}}$ 的 Markov 过程计算得到的概率. 证明 $\mathrm{P}^{\top}$ 是在如下意义下 P 的逆: 对每一 $n \in \mathbb{N}$, $0 = t_0 < t_1 < \cdots < t_n$ 和 $(j_0, \cdots, j_n) \in \mathbb{S}^{n+1}$, 有

$$\mathrm{P}^{\top}(X(t_m) = j_m, 0 \leqslant m \leqslant n) = \mathrm{P}(X(t_n - t_m) = j_m, 0 \leqslant m \leqslant n).$$

(c) 定义 $\boldsymbol{P}^{\top}$ 使得

$$(\boldsymbol{P}^{\top})_{ii} = 0 \text{ 且 } (\boldsymbol{P}^{\top})_{ij} = \frac{(\hat{\boldsymbol{\pi}})_j (\boldsymbol{P})_{ji}}{R_i (\hat{\boldsymbol{\pi}})_i}, \quad i \neq j.$$

证明 $\boldsymbol{P}^{\top}$ 也是转移概率矩阵, 并验证 $\boldsymbol{Q}^{\top} = \boldsymbol{R}(\boldsymbol{P}^{\top} - \boldsymbol{I})$.

4.5.5. 取 $\mathbb{S} = \mathbb{N}$, 并根据 j 是否等于 $i+1$, 定义 $(\boldsymbol{P})_{ij}$ 等于 1 或 0. 给定一个严格正的速率集 $\mathcal{R}$, 证明无论初始分布是什么, 如果 $\sum_{i \in \mathbb{N}} R_i^{-1} < \infty$, 则由 $\mathcal{R}$ 和 $\boldsymbol{P}$ 决定的 Markov 过程以概率 1 爆炸; 如果 $\sum_{i \in \mathbb{N}} R_i^{-1} = \infty$, 则不爆炸.

4.5.6. 这是一个关于爆炸的更有兴趣的例子. 取 $\mathbb{S}=\mathbb{Z}^3$, 设 (参看习题 3.3.5) $\boldsymbol{P}$ 是在 $\mathbb{Z}^3$ 上对称的、最近邻的随机游动的转移概率矩阵. 给定一个满足 $\sum_{\boldsymbol{k}\in\mathbb{N}} R_{\boldsymbol{k}}^{-1}<\infty$ 的正速率集 $\mathcal{R}$. 证明, 从任意 $\boldsymbol{k}\in\mathbb{Z}^3$ 出发, 当 $(\boldsymbol{Q})_{\boldsymbol{k}\boldsymbol{\ell}}=R_{\boldsymbol{k}}((\boldsymbol{P})_{\boldsymbol{k}\boldsymbol{\ell}}-\delta_{\boldsymbol{k},\boldsymbol{\ell}})$ 时, 以概率 1 发生爆炸.

提示: 应用定理 4.3.6 的第二部分的结论于如下形式的函数:

$$\sum_{\boldsymbol{\ell}\in\mathbb{Z}^3}\frac{1}{R_{\boldsymbol{\ell}}}\left(\alpha^2+\sum_{i=1}^{3}((\boldsymbol{k})_i-(\boldsymbol{\ell})_i)^2\right)^{-\frac{1}{2}},$$

并利用习题 3.3.5 中的计算方法.

4.5.7. 即使 $\mathbb{S}$ 是有限的, 要写出 (4.2.7) 式解的明确表达式也是很难的, 常常是不可能的. 虽然如此, 一些线性代数方法经常起着很好的作用. 在这一习题中, 设 $\boldsymbol{Q}$ 是状态空间 $\mathbb{S}$ 上的一个 Q-矩阵, 且对应的从任何一点出发的 Markov 过程都存在 (也即不爆炸).

(a) 设 $\boldsymbol{Q}$ 的特征值 $\alpha\in\mathbb{C}$, 如果 $\boldsymbol{u}\in\mathbb{C}^{\mathbb{S}}$ 是对应的关于 $\boldsymbol{Q}$ 的一个有界非零右特征向量, 证明 α 的实部小于或等于 0.

(b) 假设 $N=\#\mathbb{S}<\infty$, 并且 $\boldsymbol{Q}$ 拥有一个由线性独立、与特征值 $\alpha_1,\cdots,\alpha_N$ 对应的右特征向量 $\boldsymbol{u}_1,\cdots,\boldsymbol{u}_N\in\mathbb{C}^{\mathbb{S}}$ 组成的完全集. 令 $\boldsymbol{U}$ 为一个矩阵, 其第 m 列是 $\boldsymbol{u}_m$. 证明 $e^{t\boldsymbol{Q}}=\boldsymbol{U}\boldsymbol{\Lambda}(t)\boldsymbol{U}^{-1}$, 其中 $\boldsymbol{\Lambda}(t)$ 是对角矩阵, 它的第 m 个对角元素为 $e^{t\alpha_m}$.

4.5.8. 本题是习题 3.3.7 的结论在连续情形时的类比. 假设 i 是 $\boldsymbol{Q}$-常返的, 且 $R_i>0$, 对 $j\in\mathbb{S}$ 令

$$\hat{\mu}_j=\mathrm{E}\left[\int_0^{\sigma_i}\mathbf{1}_{\{j\}}(X(t))dt\middle|X(0)=i\right]\in[0,\infty],$$

假设 $\hat{\boldsymbol{\mu}}$ 是一个行向量, 对每个 $j\in\mathbb{S}$, $(\hat{\boldsymbol{\mu}})_j=\hat{\mu}_j$.

(a) 证明 $\hat{\mu}_i=\dfrac{1}{R_i}$, 对所有的 $j\in\mathbb{S}$, 有 $\hat{\mu}_j<\infty$, 且 $\hat{\mu}_j>0$ 当且仅当 $i\overset{\boldsymbol{Q}}{\leftrightarrow}j$.

(b) 证明对所有的 $t>0$, $\hat{\boldsymbol{\mu}}=\hat{\boldsymbol{\mu}}\boldsymbol{P}(t)$.

提示: 验证

$$(\hat{\boldsymbol{\mu}}\boldsymbol{P}(s))_j=\mathrm{E}\left[\int_s^{s+\sigma_i}\mathbf{1}_{\{j\}}(X(t))dt\middle|X(0)=i\right]$$

和

$$\mathrm{E}\left[\int_{\sigma_i}^{s+\sigma_i}\mathbf{1}_{\{j\}}(X(t))dt|X(0)=i\right]=\mathrm{E}\left[\int_0^{s}\mathbf{1}_{\{j\}}(X(t))dt\middle|X(0)=i\right].$$

(c) 特别地, 如果 i 是 $\boldsymbol{Q}$-正常返的, $C=\{j: i \overset{\boldsymbol{Q}}{\leftrightarrow} j\}$. 证明 $\hat{\boldsymbol{\mu}}=(\sum_j \hat{\mu}_j)\hat{\boldsymbol{\pi}}^C$. 等价地,

$$(\hat{\boldsymbol{\pi}}^C)_j=\frac{\mathrm{E}\left[\int_0^{\sigma_i}\mathbf{1}_{\{j\}}(X(t))dt|X(0)=i\right]}{\mathrm{E}[\sigma_i|X(0)=i]}.$$

4.5.9. 我们已经对连续时间给出了类似于习题 3.3.7 的结论, 现在我们对连续时间给出类似于习题 3.3.8 的结论. 为此, 假设 i 是 $\mathbb{S}$ 上的一个 $\boldsymbol{Q}$-常返状态且满足 $R_i>0$, 测度 $\hat{\boldsymbol{\mu}}$ 的定义同习题 4.5.8. 假设 $\hat{\boldsymbol{\nu}}\in[0,\infty)^{\mathbb{S}}$, 它满足条件: $(\hat{\boldsymbol{\nu}})_i>0$, $(\hat{\boldsymbol{\nu}})_j=0$ 除非 $\mathrm{P}(\sigma_j<\infty|X(0)=i)=1$, 并且 $\hat{\boldsymbol{\nu}}\boldsymbol{Q}=\mathbf{0}$, 后者的含义是 $R_j(\hat{\boldsymbol{\nu}})_j=\sum_{k\neq j}(\hat{\boldsymbol{\nu}})_k\boldsymbol{Q}_{kj}$, $j\in\mathbb{S}$.[6] 我们的目的就是要证明 $\hat{\boldsymbol{\nu}}=R_i(\hat{\boldsymbol{\nu}})_i\hat{\boldsymbol{\mu}}$. 等价地,

$$(\hat{\boldsymbol{\nu}})_j=R_i(\hat{\boldsymbol{\nu}})_i\mathrm{E}\left[\int_0^{\sigma_i}\mathbf{1}_{\{j\}}(X(t))dt\middle|X(0)=i\right],\qquad j\in\mathbb{S}.$$

特别地, 利用习题 4.5.8, 上式意味着对所有的 $t>0$ 都有 $\hat{\boldsymbol{\nu}}=\hat{\boldsymbol{\nu}}\boldsymbol{P}(t)$.

下面, 选择一个转移概率 $\boldsymbol{P}$ 使得它的对角元素为零且 $\boldsymbol{Q}=\boldsymbol{R}(\boldsymbol{P}-\boldsymbol{I})$, 其中 $\boldsymbol{R}$ 是一个对角矩阵, 它的第 j 个对角元素为 R_j, $j\in\mathbb{S}$.

(a) 证明 i 是 $\boldsymbol{P}$-常返的.

(b) 定义一个行向量 $\boldsymbol{\nu}$ 使得 $(\boldsymbol{\nu})_j=\dfrac{R_j(\hat{\boldsymbol{\nu}})_j}{R_i(\hat{\boldsymbol{\nu}})_i}$, 证明 $\boldsymbol{\nu}=\boldsymbol{\nu}\boldsymbol{P}$.

(c) 利用 (b) 和习题 3.3.7, 证明

$$(\boldsymbol{\nu})_j=\mathrm{E}\left[\sum_{m=0}^{\rho_i-1}\mathbf{1}_{\{j\}}(X_m)\middle|X_0=i\right],$$

其中 $\{X_n:n\geqslant 0\}$ 是一个 Markov 链, 它的转移概率矩阵为 $\boldsymbol{P}$.

(d) 证明

$$R_j\mathrm{E}\left[\int_0^{\sigma_i}\mathbf{1}_{\{j\}}(X(t))dt\middle|X(0)=i\right]=\mathrm{E}\left[\sum_{m=0}^{\rho_i-1}\mathbf{1}_{\{j\}}(X_m)\middle|X_0=i\right],$$

然后结合 (c), 得到想要的结论.

4.5.10. 按照习题 3.3.9 中的思路, 在推论 4.4.10 的假设下, 证明下列

[6] 这里条件 $\hat{\boldsymbol{\nu}}\boldsymbol{Q}=\mathbf{0}$ 需要补充的原因是因为 $\boldsymbol{Q}$ 有负的对角元素, 会因可能的无穷导致不确定.

连续情形时的逐点遍历定理:

$$\mathrm{P}(\lim_{T\to\infty} \|\boldsymbol{L}_T - \hat{\boldsymbol{\pi}}^C\|_v = 0) = 1,$$

其中 $C = \{i : j \overset{\boldsymbol{Q}}{\leftrightarrow} i\}$, $\boldsymbol{L}_T$ 是经验测度, 其定义为

$$(\boldsymbol{L}_T)_i = \frac{1}{T}\int_0^T \mathbf{1}_{\{i\}}(X(t))dt.$$

此外, 按照习题 3.3.11 中的思路, 证明: 对任意的初始分布,

$$\mathrm{P}\left(\lim_{T\to\infty} \frac{1}{T}\int_0^T \mathbf{1}_{\{j\}}(X(t))dt = 0\right) = 1,$$

其中 j 不是 $\boldsymbol{Q}$-正常返的.

4.5.11. 给定一个 Markov 过程 $\{X(t) : t \geqslant 0\}$, 它的 Q-矩阵为 $\boldsymbol{Q}$. 我们可用因子 M 加速 $t \rightsquigarrow X(t)$ 中的时钟, 产生一个以 $M\boldsymbol{Q}$ 为 Q-矩阵的 Markov 过程. 也就是说, $\{X(Mt) : t \geqslant 0\}$ 是一个 Markov 过程, 它的 Q-矩阵为 $M\boldsymbol{Q}$. 这一习题的目的是给出如何实现关于可变速率的类似做法, 这一方法在文献中被称为**随机时间变换**. 确切地说, 假设 $\boldsymbol{P}$ 是一个 $\mathbb{S}$ 上的转移概率矩阵, 对每个 i 满足 $(\boldsymbol{P})_{ii} = 0$, 选取并固定 $i \in \mathbb{S}$, 令 $\{X_n : n \geqslant 0\}$ 是一个从 i 出发的 Markov 链, 其转移概率矩阵为 $\boldsymbol{P}$, $\{N(t) : t \geqslant 0\}$ 是一个简单 Poisson 过程, 它与 $\{X_n : n \geqslant 0\}$ 独立, 并令 $X^0(t) = X_{N(t)}$, $t \geqslant 0$. 特别地, $\{X^0(t) : t \geqslant 0\}$ 是一个从 i 出发、以 $(\boldsymbol{P} - \boldsymbol{I})$ 为 Q-矩阵的 Markov 过程. 最后, 令 $\mathcal{R}$ 是一个正速率集并取 $\boldsymbol{Q} = \boldsymbol{R}(\boldsymbol{P} - \boldsymbol{I})$.

(a) 定义

$$A(t) = \int_0^t \frac{1}{R_{X^0(\tau)}}d\tau, \qquad t \in [0, \infty),$$

注意到 $t \rightsquigarrow A(t)$ 是严格单调增的, 记 $A(\infty) = \lim_{t\nearrow\infty} A(t) \in (0, \infty]$, 用 $s \in [0, A(\infty)) \mapsto A^{-1}(s) \in [0, \infty)$ 表示 $t \rightsquigarrow A(t)$ 的逆. 证明

$$A^{-1}(s) = \int_0^s R_{X^0(A^{-1}(\sigma))}d\sigma, \qquad s \in [0, A(\infty)).$$

(b) 记 $X(s) = X^0(A^{-1}(s))$, $s \in [0, A(\infty))$. 定义 $J_0^0 = 0 = J_0$, 对 $n \geqslant 1$, 令 J_n^0 和 J_n 分别是 $t \rightsquigarrow X^0(t)$ 和 $s \rightsquigarrow X(s)$ 的第 n 次跳跃

的时间. 注意 $J_n = A(J_n^0)$. 证明, 对每个 $n \geqslant 1$, $s > 0$ 和 $j \in \mathbb{S}$, 在 $\{J_{n-1} < \infty\}$上

$$\begin{aligned} &\mathrm{P}(J_n - J_{n-1} > s, X(J_n) = j | X(\sigma), \sigma \in [0, J_n)) \\ = &e^{-tR_{X(J_{n-1})}}(\boldsymbol{P})_{X(J_{n-1})j}. \end{aligned}$$

(c) 证明从 i 出发的 Q-矩阵为 $\boldsymbol{Q}$ 的 Markov 过程爆炸的时间与 $A(\infty)$ 是同分布的. 特别地, 如果以概率 1 成立 $A(\infty) = \infty$, 试利用 (b) 证明 $\{X(s) : s \geqslant 0\}$ 是从 i 出发、以 $\boldsymbol{Q}$ 为 Q-矩阵的 Markov 过程.

(d) 作为上述结果的一个推论, 证明: 如果相应于 $\boldsymbol{Q}$ 的 Markov 过程不爆炸, 那么, 相应于 $\boldsymbol{Q}'$ 的 Markov 过程也不爆炸, 其中 $(\boldsymbol{Q}')_{ij} = \alpha_i(\boldsymbol{Q})_{ij}$, 而 $\{\alpha_i : i \in \mathbb{S}\}$ 是一个 $(0, \infty)$ 的有界子集.

第五章　可逆 Markov 过程

这一章研究一类 Markov 过程, 它们具有一个初始分布使得关于此初始分布的过程是*可逆的*, 也就是说, 对每个时间区间, 过程向前运动时的分布与它向后运动时的分布是相同的. 即对任意的 $n \geqslant 1$ 和 $(i_0, \cdots, i_n) \in E^{n+1}$, 在离散时间情形,

$$\mathrm{P}(X_m = i_m, 0 \leqslant m \leqslant n) = \mathrm{P}(X_{n-m} = i_m, 0 \leqslant m \leqslant n); \tag{5.0.1}$$

而在连续时间情形,

$$\mathrm{P}(X(t_m) = i_m, 0 \leqslant m \leqslant n) = \mathrm{P}(X(t_n - t_m) = i_m, 0 \leqslant m \leqslant n) \tag{5.0.2}$$

对任意的 $0 = t_0 < \cdots < t_n$ 成立. 注意这种过程的初始分布必然是平稳的. 确实, 对于离散时间情形或连续时间情形, 我们有

$$\mathrm{P}(X_0 = i, X_n = j) = \mathrm{P}(X_n = i, X_0 = j)$$

或

$$\mathrm{P}(X(0) = i, X(t) = j) = \mathrm{P}(X(t) = i, X(0) = j)\ .$$

对 j 求和, 从上式即得平稳性. 事实上, 前面的讨论所揭示的是, 可逆性表明 (X_0, X_n) 或 $(X(0), X(t))$ (依赖于具体参数) 的联合分布与 (X_n, X_0) 或 $(X(t), X(0))$ 的联合分布是相同的. 而与之相对照的平稳性仅仅表示它们第一个分量的边际分布相同.

鉴于前面的讨论, 我们可以猜想, 与一般的平稳过程相比, 可逆的 Markov 过程具有更好的遍历性质, 在这一章中, 我们将会考察其特有的某些性质.

5.1 可逆 Markov 链

这一节中, 我们将要讨论不可约的可逆 Markov 链. 因为这样一个 Markov 链的初始分布是平稳的, 我们知道 (参看定理 3.2.10) 这个 Markov 链必然是正常返的, 同时, 它的初始分布一定是概率向量 (参看 (3.2.9)) $\boldsymbol{\pi} = \boldsymbol{\pi}^{\mathbb{S}}$, 其第 i 个分量是 $(\boldsymbol{\pi})_i \equiv \mathrm{E}[\rho_i|X_0 = i]^{-1}$. 因此, 如果 $\boldsymbol{P}$ 是转移概率矩阵, 在 (5.0.1) 式中取 $n = 1$, 就有

$$(\boldsymbol{\pi})_i(\boldsymbol{P})_{ij} = \mathrm{P}(X_0 = i, X_1 = j) = \mathrm{P}(X_0 = j, X_1 = i) = (\boldsymbol{\pi})_j(\boldsymbol{P})_{ji}.$$

就是说, $\boldsymbol{P}$ 满足[1] 如下的**逐一平衡条件**

$$(\boldsymbol{\pi})_i(\boldsymbol{P})_{ij} = (\boldsymbol{\pi})_j(\boldsymbol{P})_{ji}. \tag{5.1.1}$$

相反地, 从 (5.1.1) 式可推得可逆性. 事实上, 对 $n \geqslant 1$ 进行归纳得到

$$(\boldsymbol{\pi})_{i_0}(\boldsymbol{P})_{i_0 i_1} \cdots (\boldsymbol{P})_{i_{n-1} i_n} = (\boldsymbol{\pi})_{i_n}(\boldsymbol{P})_{i_n i_{n-1}} \cdots (\boldsymbol{P})_{i_1 i_0}.$$

上式等价于 (5.0.1) 式.

5.1.1. 从不变性到可逆性: 我们已经知道, 可逆性可推得不变性, 而反过来并不成立. 另一方面, 有两种经典的方法可以把一个不可约的带平稳分布 $\boldsymbol{\pi}$ 的转移概率 $\boldsymbol{P}$ 转换成一个转移概率, 使得 $\boldsymbol{\pi}$ 关于它是可逆的. 即定义 $\boldsymbol{P}$ 的**伴随概率矩阵** $\boldsymbol{P}^\top$ 使得

$$(\boldsymbol{P}^\top)_{ij} = \frac{(\boldsymbol{\pi})_j(\boldsymbol{P})_{ij}}{(\boldsymbol{\pi})_i}. \tag{5.1.2}$$

显然, 对于 $\boldsymbol{P}, \boldsymbol{\pi}$ 是可逆的当且仅当 $\boldsymbol{P} = \boldsymbol{P}^\top$. 更一般地, 因为 $\boldsymbol{\pi}\boldsymbol{P} = \boldsymbol{\pi}$, 故 $\boldsymbol{P}^\top$ 也是一个转移概率. 我们也能容易地得到

$$\frac{\boldsymbol{P} + \boldsymbol{P}^\top}{2} \quad 和 \quad \boldsymbol{P}^\top\boldsymbol{P} \tag{5.1.3}$$

都是转移概率且它们关于 $\boldsymbol{\pi}$ 都是可逆的. 从下面的习题 5.6.9 中的解释可以知道, 这些构造各有它独特的优点.

5.1.2. 二次平均度量: 这里我们将用欧氏范数代替一致范数 $\|f\|_u$ 去度量函数的大小, 这样做的理由后面将会变得越来越清晰. 也就是说, 我们定义范数

$$\|f\|_{2,\boldsymbol{\pi}} \equiv \sqrt{\langle |f|^2 \rangle_{\boldsymbol{\pi}}}, \quad 其中\ \langle g \rangle_{\boldsymbol{\pi}} \equiv \sum_i g(i)(\boldsymbol{\pi})_i \tag{5.1.4}$$

[1] 做过习题 2.4.1 的读者应该认识到下面的条件等同于 $\boldsymbol{P} = \boldsymbol{P}^\top$. 特别地, 如果知道了习题 2.4.1 的结论, 那么就没有必要看接下去的讨论了.

为 g 关于 $\boldsymbol{\pi}$ 的数学期望值. 因为, 对每个 $i \in \mathbb{S}$, $(\boldsymbol{\pi})_i > 0$, 故

$$||f||_{2,\boldsymbol{\pi}} = 0 \Longleftrightarrow f \equiv 0.$$

此外, 如果我们定义内积 $\langle f, g\rangle_{\boldsymbol{\pi}}$ 为 $\langle fg\rangle_{\boldsymbol{\pi}}$, 则对任意的 $t > 0$, 有

$$0 \leqslant ||tf \pm t^{-1}g||_{2,\boldsymbol{\pi}}^2 = t^2||f||_{2,\boldsymbol{\pi}}^2 \pm 2\langle f, g\rangle_{\boldsymbol{\pi}} + t^{-2}||g||_{2,\boldsymbol{\pi}}^2.$$

因此对所有的 $t > 0$ 有 $|\langle f, g\rangle_{\boldsymbol{\pi}}| \leqslant t^2||f||_{2,\boldsymbol{\pi}}^2 + t^{-2}||g||_{2,\boldsymbol{\pi}}^2$. 为了得到最好的估计, 我们关于 $t > 0$ 极小化上式的右边. 当 $f = 0$ 或 $g = 0$ 时, 令 $t \to \infty$ 或 $t \to 0$, 得到 $\langle f, g\rangle_{\boldsymbol{\pi}} = 0$. 如果 f 和 g 均不为零, 取 $t = \left(\frac{||g||_{2,\boldsymbol{\pi}}}{||f||_{2,\boldsymbol{\pi}}}\right)^{\frac{1}{4}}$, 可得到最佳结果. 因此, 无论何种情形, 我们都有下面的 *Schwarz* 不等式

$$|\langle f, g\rangle_{\boldsymbol{\pi}}| \leqslant ||f||_{2,\boldsymbol{\pi}}||g||_{2,\boldsymbol{\pi}}. \tag{5.1.5}$$

由这一不等式, 我们有

$$||f + g||_{2,\boldsymbol{\pi}}^2 = ||f||_{2,\boldsymbol{\pi}}^2 + 2\langle f, g\rangle_{\boldsymbol{\pi}} + ||g||_{2,\boldsymbol{\pi}}^2 \leqslant (||f||_{2,\boldsymbol{\pi}} + ||g||_{2,\boldsymbol{\pi}})^2 .$$

这就是说, 我们有三角不等式:

$$||f + g||_{2,\boldsymbol{\pi}} \leqslant ||f||_{2,\boldsymbol{\pi}} + ||g||_{2,\boldsymbol{\pi}}. \tag{5.1.6}$$

因此, 如果 $L^2(\boldsymbol{\pi})$ 表示由满足 $||f||_{2,\boldsymbol{\pi}} < \infty$ 的函数 f 组成的空间, 则 $L^2(\boldsymbol{\pi})$ 是一个线性空间且 $(f, g) \rightsquigarrow ||f - g||_{2,\boldsymbol{\pi}}$ 为此空间上的一个度量. 这个度量空间是完备的. 事实上, 如果 $\lim\limits_{m\to\infty} \sup\limits_{n>m} ||f_n - f_m||_{2,\boldsymbol{\pi}} = 0$, 则对每个 $i \in \mathbb{S}$, $\{f_n(i) : n \geqslant 0\}$ 在 $\mathbb{R}$ 上是 Cauchy 收敛的. 因此, 存在一个极限函数 f 使得对每个 $i \in \mathbb{S}$ 有 $f_n(i) \to f(i)$. 进一步, 由 Fatou 引理得到

$$||f - f_m||_{2,\boldsymbol{\pi}} \leqslant \liminf_{n\to\infty} ||f_n - f_m||_{2,\boldsymbol{\pi}} \to 0, \quad m \to \infty.$$

在这一章中我们将使用记号

$$\boldsymbol{P}f(i) = \sum_{j\in\mathbb{S}} f(j)(\boldsymbol{P})_{ij} = (\boldsymbol{P}\boldsymbol{f})_i,$$

其中 $\boldsymbol{f}$ 是由 f 决定的列向量. 当 f 有界时, 易知 $\boldsymbol{P}f(i)$ 是有定义的且有 $||\boldsymbol{P}f||_u \leqslant ||f||_u$, 其中 $||g||_{\boldsymbol{u}} \equiv \sup\limits_{i\in\mathbb{S}} |g(i)| = ||\boldsymbol{g}||_{\boldsymbol{u}}$, $\boldsymbol{g}$ 是由 g 决定的

列向量. 我们接下来证明, 甚至当 $f \in L^2(\boldsymbol{\pi})$ 时, $\boldsymbol{P}f(i)$ 对每个 $i \in \mathbb{S}$ 是有定义的且 $\boldsymbol{P}$ 是 $L^2(\boldsymbol{\pi})$ 上的一个压缩映射: $\|\boldsymbol{P}f\|_{2,\boldsymbol{\pi}} \leqslant \|f\|_{2,\boldsymbol{\pi}}$. 为了验证 $\boldsymbol{P}f(i)$ 是有定义的, 我们证明级数 $\sum_{j\in\mathbb{S}} f(j)(\boldsymbol{P})_{ij}$ 是绝对收敛的. 利用习题 1.3.1 中的 Schwarz 不等式可知

$$\sum_{j\in\mathbb{S}} |f(j)|(\boldsymbol{P})_{ij} \leqslant \|f\|_{2,\boldsymbol{\pi}} \left(\sum_{j\in\mathbb{S}} \frac{(\boldsymbol{P})_{ij}^2}{(\boldsymbol{\pi})_j}\right)^{\frac{1}{2}},$$

同时, 由 (5.1.1) 式可知

$$\sum_{j\in\mathbb{S}} \frac{(\boldsymbol{P})_{ij}^2}{(\boldsymbol{\pi})_j} = \frac{1}{(\boldsymbol{\pi})_i}\sum_{j\in\mathbb{S}}(\boldsymbol{P})_{ij}(\boldsymbol{P})_{ji} = \frac{(\boldsymbol{P}^2)_{ii}}{(\boldsymbol{\pi}_i)} < \infty.$$

现在证明 $\|\boldsymbol{P}f\|_{2,\boldsymbol{\pi}} \leqslant \|f\|_{2,\boldsymbol{\pi}}$. 利用下面的习题 5.6.2 和 $\sum_j(\boldsymbol{P})_{ij} = 1$ 可知, 对每个 i 有 $(\boldsymbol{P}f(i))^2 \leqslant \boldsymbol{P}f^2(i)$. 因为 $\boldsymbol{\pi}$ 是 $\boldsymbol{P}$-平稳的, 所以

$$\|\boldsymbol{P}f\|_{2,\boldsymbol{\pi}} \leqslant \|f\|_{2,\boldsymbol{\pi}}. \tag{5.1.7}$$

(5.1.7) 式的一个重要推论是, 一般而言 (参看 (2.2.4)), 对所有的 $f \in L^2(\boldsymbol{\pi})$ 有

$$\lim_{n\to\infty} \|\boldsymbol{A}_n f - \langle f\rangle_{\boldsymbol{\pi}}\|_{2,\boldsymbol{\pi}} = 0 \tag{5.1.8}$$

且

$$\boldsymbol{P}\text{ 是非周期的} \implies \lim_{n\to\infty} \|\boldsymbol{P}^n f - \langle f\rangle_{\boldsymbol{\pi}}\|_{2,\boldsymbol{\pi}} = 0. \tag{5.1.9}$$

事实上, 首先注意到, 由 (3.2.11), (3.2.15) 式和 Lebesgue 控制收敛定理可知, 当 f 在一个有限集外为零时, 则上面两式显然成立. 如果 $\{F_N : N \geqslant 1\}$ 是 $\mathbb{S}$ 的一个有限集的穷举 (即对每一 N, F_N 是 $\mathbb{S}$ 的一个非空有限子集, $F_N \subseteq F_{N+1}$, 且 $\mathbb{S} = \bigcup_N F_N$), 且对 $f \in L^2(\boldsymbol{\pi})$, 记 $f_N \equiv \mathbf{1}_{F_N} f$, 则对每个 $N \in \mathbb{Z}^+$ 有

$$\begin{aligned}
\|\boldsymbol{A}_n f - \langle f\rangle_{\boldsymbol{\pi}}\|_{2,\boldsymbol{\pi}} \leqslant & \|\boldsymbol{A}_n(f - f_N)\|_{2,\boldsymbol{\pi}} + \|\boldsymbol{A}_n f_N - \langle f_N\rangle_{\boldsymbol{\pi}}\|_{2,\boldsymbol{\pi}} + \\
& \langle |f_N - f|\rangle_{\boldsymbol{\pi}} \\
\leqslant & 2\|f - f_N\|_{2,\boldsymbol{\pi}} + \|\boldsymbol{A}_n f_N - \langle f_N\rangle_{\boldsymbol{\pi}}\|_{2,\boldsymbol{\pi}},
\end{aligned}$$

其中第二个不等式利用了 $\|\boldsymbol{A}_n g\|_{2,\boldsymbol{\pi}} \leqslant \|g\|_{2,\boldsymbol{\pi}}$, 此式可从 (5.1.7) 式直接得到. 因此, 对每个 N,

$$\limsup_{n\to\infty} \|\boldsymbol{A}_n f - \langle f\rangle_{\boldsymbol{\pi}}\|_{2,\boldsymbol{\pi}} \leqslant 2\|f - f_N\|_{2,\boldsymbol{\pi}}.$$

令 $N \to \infty$, 即得 (5.1.8) 式. (5.1.9) 式可类似得到. 后面的习题 5.6.6 指出, 上面这些结论即使在非可逆情形也同样成立.

最后, 由 (5.1.1) 式得到

$$\langle g, \boldsymbol{P}f\rangle_{\boldsymbol{\pi}} = \sum_{(i,j)}(\boldsymbol{\pi})_i g(i)(\boldsymbol{P})_{ij}f(j) = \sum_{(i,j)}(\boldsymbol{\pi})_j f(i)(\boldsymbol{P})_{ji}g(j) = \langle \boldsymbol{P}g, f\rangle_{\boldsymbol{\pi}}.$$

换句话说, $\boldsymbol{P}$ 在 $L^2(\boldsymbol{\pi})$ 上是对称的, 也即对 $(f,g) \in (L^2(\boldsymbol{\pi}))^2$,

$$\langle g, \boldsymbol{P}f\rangle_{\boldsymbol{\pi}} = \langle \boldsymbol{P}g, f\rangle_{\boldsymbol{\pi}}\ . \tag{5.1.10}$$

5.1.3. 谱隙: 结合 (5.1.7) 和 (5.1.10) 得知在 Hilbert 空间[2] 上的一个 $\boldsymbol{P}$ 自伴随压缩是 $L^2(\boldsymbol{\pi})$. 考虑到读者可能不习惯这些抽象的概念, 我们考察 $\mathbb{S} = \{1, \cdots, N\}$ 的情形. 这时函数 f: $\mathbb{S} \to \mathbb{R}$ 的空间可以看做是 $\mathbb{R}^N$. (在 §2.1.3 中给出函数与列向量之间的关系时, 我们就已经有过这样视它们为同一的做法.) 在有了这样的认识后, $L^2(\boldsymbol{\pi})$ 上的内积就变成了内积 $\sum_1^N(\boldsymbol{\pi})_i(\boldsymbol{v})_i(\boldsymbol{w})_i$, 其中 $\boldsymbol{v}$, $\boldsymbol{w} \in \mathbb{R}^N$ 为列向量. 因此 (5.1.7) 式表示矩阵 $\boldsymbol{P}$ 关于这个内积在 $\mathbb{R}^N$ 上起了一个对称的压缩映射的作用. 如果 $\tilde{\boldsymbol{P}} \equiv \boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{P}\boldsymbol{\Pi}^{-\frac{1}{2}}$, 其中 $\boldsymbol{\Pi}$ 是一个对角矩阵, 它的第 i 个对角元素为 $(\boldsymbol{\pi})_i$, 则由 (5.1.1) 式可知, $\tilde{\boldsymbol{P}}$ 关于 $\mathbb{R}^N$ 上的标准内积 $(\boldsymbol{v}, \boldsymbol{w})_{\mathbb{R}^N} \equiv \sum_1^N(\boldsymbol{v})_i(\boldsymbol{w})_i$ 是对称的. 进一步, 由 (5.1.7) 式可知

$$\begin{aligned}||\tilde{\boldsymbol{P}}\boldsymbol{f}||_{\mathbb{R}^N}^2 &\equiv (\tilde{\boldsymbol{P}}\boldsymbol{f}, \tilde{\boldsymbol{P}}\boldsymbol{f})_{\mathbb{R}^N} = \sum_i(\boldsymbol{\pi})_i(\boldsymbol{P}\boldsymbol{\Pi}^{-\frac{1}{2}}\boldsymbol{f})_i^2 \\ &= ||\boldsymbol{P}g||_{2,\boldsymbol{\pi}}^2 \leqslant ||g||_{2,\boldsymbol{\pi}}^2 = ||\boldsymbol{f}||_{\mathbb{R}^N}^2,\end{aligned}$$

其中 g 是由列向量 $\boldsymbol{\Pi}^{-\frac{1}{2}}\boldsymbol{f}$ 决定的函数. 因此, 作为 $\mathbb{R}^N$ 上的一个算子, $\tilde{\boldsymbol{P}}$ 是长度压缩. 现在, 利用 $\mathbb{R}^N$ 上的对称矩阵的标准理论可知, $\tilde{\boldsymbol{P}}$ 具有特征值 $1 \geqslant \lambda_1 \geqslant \cdots \geqslant \lambda_N \geqslant -1$, 且相应的特征向量 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_N)$ 关于 $(\cdot,\cdot)_{\mathbb{R}^N}$ 是正交的: $(\boldsymbol{e}_k, \boldsymbol{e}_\ell)_{\mathbb{R}^N} = \delta_{k,\ell}$. 进一步, 由于 $\sqrt{(\boldsymbol{\pi})_i} = \sum_{j=1}^N (\tilde{\boldsymbol{P}})_{ij}\sqrt{(\boldsymbol{\pi})_j}$, 我们知道 $\lambda_1 = 1$, 并且可以取 $(\boldsymbol{e}_1)_i = \sqrt{(\boldsymbol{\pi})_i}$. 最后, 取 $\boldsymbol{g}_\ell = (\boldsymbol{\Pi})^{-\frac{1}{2}}\boldsymbol{e}_\ell$ 并令 g_ℓ 是 $\mathbb{S}$ 上的相伴随的函数, 我们得到 $\boldsymbol{P}g_\ell = \lambda_\ell g_\ell$, $g_1 \equiv 1$ 和 $\langle g_k, g_\ell\rangle_{\boldsymbol{\pi}} = \delta_{k,l}$. 概括地说, 当 $\mathbb{S}$ 有 N 个元素

[2] Hilbert 空间是一个赋有内积的向量空间, 由此内积确定的范数所定义的度量是完备的.

时, 我们已经知道 $\boldsymbol{P}$ 在 $L^2(\boldsymbol{\pi})$ 上有特征值 $1 \geqslant \lambda_1 \geqslant \cdots \geqslant \lambda_N \geqslant -1$, 且相应的特征函数 $g_1, \cdots, g_N$ 关于 $(\cdot,\cdot)_{\boldsymbol{\pi}}$ 是正交的. 当然, 因为 $L^2(\boldsymbol{\pi})$ 是 N 维的, 利用标准正交性, 得知 g_ℓ 是线性独立的, $(g_1, \cdots, g_N)$ 是 $L^2(\boldsymbol{\pi})$ 中的一个正交基. 特别地, 对所有的 $n \geqslant 0, f \in L^2(\boldsymbol{\pi})$,

$$\boldsymbol{P}^n f - \langle f \rangle_{\boldsymbol{\pi}} = \sum_{\ell=2}^{N} \lambda_\ell^n \langle f, g_\ell \rangle_{\boldsymbol{\pi}} g_\ell \,, \tag{5.1.11}$$

因此

$$||\boldsymbol{P}^n f - \langle f \rangle_{\boldsymbol{\pi}}||_{2,\boldsymbol{\pi}}^2 = \sum_{\ell=2}^{N} \lambda_\ell^{2n} \langle f, g_\ell \rangle_{\boldsymbol{\pi}}^2 \leqslant (1-\beta)^{2n} ||f - \langle f \rangle_{\boldsymbol{\pi}}||_{2,\boldsymbol{\pi}}^2,$$

其中 $\beta = (1-\lambda_2) \wedge (1+\lambda_N)$ 是 $\{-1, 1\}$ 和 $\{\lambda_2, \cdots, \lambda_N\}$ 之间的谱隙. 换句话说, 对所有的 $n \geqslant 0, f \in L^2(\boldsymbol{\pi})$,

$$||\boldsymbol{P}^n f - \langle f \rangle_{\boldsymbol{\pi}}||_{2,\boldsymbol{\pi}} \leqslant (1-\beta)^n ||f - \langle f \rangle_{\boldsymbol{\pi}}||_{L^2(\boldsymbol{\pi})} \,. \tag{5.1.12}$$

当 $\mathbb{S}$ 不是有限时, 一般而言不大可能找到一个关于 $\boldsymbol{P}$ 的特征函数的正交基. 与上述讨论最接近的一个逼近方法需要一个关于 Hilbert 空间上的有界对称算子的著名结果, 即称为谱定理的结果 (参看 [7] 中的 §107). 虽然如此, 但是我们最感兴趣的是估计 (5.1.12), 即使没有谱定理的帮助我们也能估计. 确切地说, 注意到, 当

$$\beta \equiv 1 - \sup\{||\boldsymbol{P}f - \langle f \rangle_{\boldsymbol{\pi}}||_{2,\boldsymbol{\pi}} : f \in L^2(\boldsymbol{\pi}), ||f||_{2,\boldsymbol{\pi}} = 1\} \tag{5.1.13}$$

时, (5.1.12) 式成立. 为了证明这个结论, 首先注意当 β 由 (5.1.13) 式给定时,

$$\begin{aligned}||\boldsymbol{P}f - \langle f \rangle_{\boldsymbol{\pi}}||_{2,\boldsymbol{\pi}} &= ||f||_{2,\boldsymbol{\pi}} \left\| \boldsymbol{P}\left(\frac{f}{||f||_{2,\boldsymbol{\pi}}}\right) - \left\langle \frac{f}{||f||_{2,\boldsymbol{\pi}}} \right\rangle_{\boldsymbol{\pi}} \right\|_{2,\boldsymbol{\pi}} \\ &\leqslant (1-\beta)||f||_{2,\boldsymbol{\pi}}, \quad \text{其中 } f \in L^2(\boldsymbol{\pi}) \setminus \{\mathbf{0}\},\end{aligned}$$

且当 $f = \mathbf{0}$ 时, 显然有 $||\boldsymbol{P}f - \langle f \rangle_{\boldsymbol{\pi}}||_{2,\boldsymbol{\pi}} \leqslant (1-\beta)||f||_{2,\boldsymbol{\pi}}$. 其次由 $\langle \boldsymbol{P}^n f \rangle_{\boldsymbol{\pi}} = \langle f \rangle_{\boldsymbol{\pi}}, n \geqslant 0$, 可知

$$\begin{aligned}||\boldsymbol{P}^{n+1} f - \langle f \rangle_{\boldsymbol{\pi}}||_{2,\boldsymbol{\pi}} &= ||\boldsymbol{P}(\boldsymbol{P}^n f - \langle \boldsymbol{P}^n f \rangle_{\boldsymbol{\pi}})||_{2,\boldsymbol{\pi}} \\ &\leqslant (1-\beta)||\boldsymbol{P}^n f - \langle \boldsymbol{P}^n f \rangle_{\boldsymbol{\pi}}||_{2,\boldsymbol{\pi}} \\ &= (1-\beta)||\boldsymbol{P}^n f - \langle f \rangle_{\boldsymbol{\pi}}||_{2,\boldsymbol{\pi}}.\end{aligned}$$

于是对 n 进行归纳得到, 对所有的 $n \geqslant 0$ 和 $f \in L^2(\boldsymbol{\pi})$, $\|\boldsymbol{P}^n f - \langle f\rangle_{\boldsymbol{\pi}}\|_{2,\boldsymbol{\pi}} \leqslant (1-\beta)^n \|f\|_{2,\boldsymbol{\pi}}$. 因此, 如果 $f \in L^2(\boldsymbol{\pi})$ 且记 $\bar{f} = f - \langle f\rangle_{\boldsymbol{\pi}}$, 则有 $\|\boldsymbol{P}^n f - \langle f\rangle_{\boldsymbol{\pi}}\|_{2,\boldsymbol{\pi}} = \|\boldsymbol{P}^n \bar{f}\|_{2,\boldsymbol{\pi}} \leqslant (1-\beta)^n \|\bar{f}\|_{2,\boldsymbol{\pi}} = (1-\beta)^n \|f - \langle f\rangle_{\boldsymbol{\pi}}\|_{2,\boldsymbol{\pi}}$. 就是说, 对由 (5.1.13) 式给定的 β, (5.1.12) 式成立. 注意当 $\mathbb{S}$ 有限时, (5.1.13) 式定义的 β 与前面一段里的 β 是一致的. 因此, 当 (5.1.11) 式不适用时, 就如何推广 (5.1.12) 式, 我们已经跨出了第一步.

5.1.4. 可逆性和周期性: 显然, 由 (5.1.13) 式定义的 β 可以取值为 0, 此时, (5.1.12) 式不能说明什么问题. 而这种情况发生的可能方式有三种: 第一种是存在一个函数 $f \in L^2(\boldsymbol{\pi})$, 具有下面的性质: $\|f\|_{2,\boldsymbol{\pi}} = 1$, $\langle f\rangle_{\boldsymbol{\pi}} = 0$ 且 $\boldsymbol{P}f = f$. 然而, 由不可约性可知, 这样的 f 是不存在的. 事实上, 因为不可约性, 我们会有 (参看 (5.1.8)) 下面的矛盾: 对所有的 $i \in \mathbb{S}$, $0 = \langle f\rangle_{\boldsymbol{\pi}} = \lim_{n\to\infty} (\boldsymbol{A}_n \boldsymbol{f})_i = f(i)$, 而对某些 $i \in \mathbb{S}$, $f(i) \neq 0$. 因此, 我们可以忽略这种情形, 因为它不可能发生. 第二种可能情形是存在一个函数 $f \in L^2(\boldsymbol{\pi})$, 满足 $\|f\|_{2,\boldsymbol{\pi}} = 1$ 和 $\boldsymbol{P}f = -f$. 事实上, 如果有这样的函数 f, 则必然有 $\langle f\rangle_{\boldsymbol{\pi}} = \langle \boldsymbol{P}f\rangle_{\boldsymbol{\pi}} = -\langle f\rangle_{\boldsymbol{\pi}}$, 故 $\langle f\rangle_{\boldsymbol{\pi}} = 0$. 因此, 我们有 $\|\boldsymbol{P}f - \langle f\rangle_{\boldsymbol{\pi}}\|_{2,\boldsymbol{\pi}} = 1$, 从而有 $\beta = 0$. 第三种可能情形是不存在关于 $\boldsymbol{P}f = -f$ 的非零解, 然而, 存在序列 $\{f_n\}_1^\infty \subseteq L^2(\boldsymbol{\pi})$, 它满足 $\|f_n\|_{2,\boldsymbol{\pi}} = 1$ 和 $\langle f_n\rangle_{\boldsymbol{\pi}} = 0$, 使得 $\|\boldsymbol{P}f_n\|_{2,\boldsymbol{\pi}}$ 趋于 1.

因为第三种可能情形的分析需要谱定理, 故我们不去讨论它. 然而, 下面的定理告诉我们第二种可能情形有一个令人欣慰的简单的概率解释. 后面的习题 5.6.7 推广这类结论到不可逆 $\boldsymbol{P}$ 的情形.

5.1.14 定理. *如果 $\boldsymbol{P}$ 是一个不可约的转移概率, 这时它的可逆初始分布存在且一定是 $\boldsymbol{\pi}$, 则 $\boldsymbol{P}$ 的周期是 1 或 2. 进而, 周期是 2 当且仅当存在一个函数 $f \in L^2(\boldsymbol{\pi}) \setminus \{\boldsymbol{0}\}$ 且 $f = -\boldsymbol{P}f$.*

证明: 设周期为 d, 我们先来证明 $d \leqslant 2$. 由不可约性, 对一切 i, $(\boldsymbol{\pi})_i > 0$. 因此从逐一平衡条件 (5.1.1) 式可推得, $(\boldsymbol{P})_{ij} > 0 \Longleftrightarrow (\boldsymbol{P})_{ji} > 0$. 特别地, 因为对每个 i 和某些 j 有 $(\boldsymbol{P})_{ij} > 0$, 故 $(\boldsymbol{P}^2)_{ii} = \sum_j (\boldsymbol{P})_{ij}(\boldsymbol{P})_{ji} > 0$, 因此 2 一定能被周期整除.

为了完成定理的证明, 首先假设 $d = 1$. 如果 $f \in L^2(\boldsymbol{\pi})$ 满足 $f = -\boldsymbol{P}f$, 则由前面讨论可知 $\langle f\rangle_{\boldsymbol{\pi}} = 0$, 进一步, 由非周期性和 (5.1.9) 式可知, 对每个 $i \in \mathbb{S}$ 有 $\lim_{n\to\infty} \boldsymbol{P}^n f(i) = \langle f\rangle_{\boldsymbol{\pi}} = 0$. 因为对任意的 $n \geqslant 0$, $f = \boldsymbol{P}^{2n} f$, 故得 $f \equiv 0$. 相反地, 如果 $d = 2$, 对 §3.2.7 中定义的 $\mathbb{S}_0, \mathbb{S}_1$, 考虑 $f = \mathbf{1}_{\mathbb{S}_0} - \mathbf{1}_{\mathbb{S}_1}$. 由 (3.2.19) 式可知, $\boldsymbol{P}f = -f$ 且 $\|f\|_{2,\boldsymbol{\pi}} = 1$. □

作为定理的直接推论, 我们可以给出不可约可逆 Markov 链的非周期性图理论的描述. 如果我们用 $\boldsymbol{P}$ 去定义一个图结构, $\mathbb{S}$ 的元素是 "顶点", i 与 j 之间的 "边" 存在当且仅当 $(\boldsymbol{P})_{ij} > 0$, 则结合定理 5.1.14 的第一部分和 § 3.2.7 中的内容可以得知, 这样的图是双分图 (即分成两个部分使得所有的边都把一个部分和另一个部分相连) 当且仅当 Markov 链不是非周期的. 定理的第二部分则说明这是可能的当且仅当存在一个函数 $f \in L^2(\boldsymbol{\pi}) \setminus \{\mathbf{0}\}$, 满足 $\boldsymbol{P}f = -f$.

5.1.5. 与变差收敛的关系: 在开始讨论计算和估计 (5.1.13) 式中的 β 的方法前, 我们先来比较一下包含在 (5.1.12) 式中的收敛结果与迄今为止已经得到的收敛结果. 首先注意到

$$||\boldsymbol{P}^n f - \langle f\rangle_{\boldsymbol{\pi}}||_{2,\boldsymbol{\pi}} \leqslant ||\boldsymbol{P}^n f - \langle f\rangle_{\boldsymbol{\pi}}||_u \leqslant \sup_i ||\boldsymbol{\delta}_i \boldsymbol{P}^n - \boldsymbol{\pi}||_v ||f||_u.$$

特别地, 由于利用定理 2.2.1, 对某个 $C < \infty$ 和 $\epsilon \in (0,1]$, 有

$$\sup_i ||\boldsymbol{\delta}_i \boldsymbol{P}^n - \boldsymbol{\pi}||_v \leqslant C(1-\epsilon)^n, \tag{$*$}$$

因此

$$||\boldsymbol{P}^n f - \langle f\rangle_{\boldsymbol{\pi}}||_{2,\boldsymbol{\pi}} \leqslant C(1-\epsilon)^n ||f||_u \ .$$

上式很像 (5.1.12). 的确, 唯一的区别是在 (5.1.12) 式的右边, $C = 1$, $||f||_{2,\boldsymbol{\pi}}$ 代替了 $||f||_u$. 因此, 我们从 $(*)$ 式可猜测 (5.1.12) 式中的 β 不小于 $(*)$ 式中的 ϵ. 在 $\mathbb{S}$ 有限这一总是要考察的情形, 存在一个函数 $g \in L^2(\boldsymbol{\pi})$, 满足 $||g||_{2,\boldsymbol{\pi}} = 1$, $\langle g\rangle_{\boldsymbol{\pi}} = 0$, 且要么 $\boldsymbol{P}g = (1-\beta)g$, 要么 $\boldsymbol{P}g = -(1-\beta)g$, 从而上述猜测很容易得到验证. 事实上, 记 $f = g\mathbf{1}_{[-R,R]}(g)$, 其中选取 $R > 0$ 使 $a \equiv \langle f, g\rangle_{\boldsymbol{\pi}} \geqslant \frac{1}{2}$. 又记 $\bar{f} = f - \langle f\rangle_{\boldsymbol{\pi}}$. 那么利用 $\bar{f} = ag + (\bar{f} - ag)$ 和 $\langle g, \bar{f} - ag\rangle_{\boldsymbol{\pi}} = 0$, 并注意到

$$\pm\langle g, \boldsymbol{P}^n(\bar{f} - ag)\rangle_{\boldsymbol{\pi}} = \langle \boldsymbol{P}^n g, \bar{f} - ag\rangle_{\boldsymbol{\pi}} = (1-\beta)^n \langle g, \bar{f} - ag\rangle_{\boldsymbol{\pi}} = 0,$$

对任意 $n \geqslant 0$, 有

$$\begin{aligned}
||\boldsymbol{P}^n \bar{f}||_{2,\boldsymbol{\pi}}^2 &= a^2(1-\beta)^{2n} \pm 2a(1-\beta)^n \langle g, \boldsymbol{P}^n(\bar{f} - ag)\rangle_{\boldsymbol{\pi}} + \\
&\quad ||\boldsymbol{P}^n(\bar{f} - ag)||_{2,\boldsymbol{\pi}}^2 \\
&\geqslant \frac{1}{4}(1-\beta)^{2n} \ .
\end{aligned}$$

另一方面, $\|P^n\bar{f}\|_{2,\pi}^2 \leqslant C^2(1-\epsilon)^{2n}\|\bar{f}\|_u^2 \leqslant (CR)^2(1-\epsilon)^{2n}$. 因此, 仅当 $\beta \geqslant \epsilon$ 时对一切 $n \geqslant 0$, 有 $\frac{1}{4}(1-\beta)^{2n} \leqslant (CR)^2(1-\epsilon)^{2n}$. 当这样的函数 g 不存在时, 如果要得到相同的结果, 我们只能求助于谱定理.

如前面所示, 关于 $\boldsymbol{\mu}\boldsymbol{P}^n$ 与 $\boldsymbol{\pi}$ 之间的变差距离的一致估计蕴涵了很像 (5.1.12) 式中的估计. 然而, 相反的方向不一定总是可能的. 为了验证什么是可以做的, 给定一个概率向量 $\boldsymbol{\mu}$ 后, 定义函数 f 使得 $f(i) = \frac{(\boldsymbol{\mu})_i}{(\boldsymbol{\pi})_i}$. 那么, 由于 $\langle f\rangle_{\boldsymbol{\pi}} = 1$ 且对所有的 $g \in L^2(\boldsymbol{\pi})$ 有 $\langle g\rangle_{\boldsymbol{\pi}}^2 \leqslant \langle g^2\rangle_{\boldsymbol{\pi}}$, 我们有

$$\begin{aligned}
\|\boldsymbol{\mu}\boldsymbol{P}^n - \boldsymbol{\pi}\|_v &= \sum_j |(\boldsymbol{\mu}\boldsymbol{P}^n)_j - (\boldsymbol{\pi})_j| = \sum_j |\langle f, \boldsymbol{P}^n \mathbf{1}_{\{j\}}\rangle_{\boldsymbol{\pi}} - (\boldsymbol{\pi})_j| \\
&= \sum_j |\langle \boldsymbol{P}^n f - \mathbf{1}, \mathbf{1}_{\{j\}}\rangle_{\boldsymbol{\pi}}| = \sum_j |\langle \boldsymbol{P}^n f - \langle f\rangle_{\boldsymbol{\pi}}, \mathbf{1}_{\{j\}}\rangle_{\boldsymbol{\pi}}| \\
&= \sum_j (\boldsymbol{\pi})_j |\boldsymbol{P}^n f(j) - \langle f\rangle_{\boldsymbol{\pi}}| \leqslant \left(\sum_j (\boldsymbol{\pi})_j |\boldsymbol{P}^n f(j) - \langle f\rangle_{\boldsymbol{\pi}}|^2\right)^{\frac{1}{2}} \\
&= \|\boldsymbol{P}^n f - \langle f\rangle_{\boldsymbol{\pi}}\|_{2,\boldsymbol{\pi}}.
\end{aligned}$$

因此, 由 (5.1.12) 有

$$\|\boldsymbol{\mu}\boldsymbol{P}^n - \boldsymbol{\pi}\|_v \leqslant \left(\sum_i \frac{(\boldsymbol{\mu})_i^2}{(\boldsymbol{\pi})_i} - 1\right)^{\frac{1}{2}} (1-\beta)^n. \tag{5.1.15}$$

在 $\mathbb{S}$ 是有限的情形, 存在 $\lambda \in (0,1)$, 使得 $(\boldsymbol{\pi})_i \geqslant \lambda$. 从 (5.1.15) 式得到 Doeblin 型估计

$$\|\boldsymbol{\mu}\boldsymbol{P}^n - \boldsymbol{\pi}\|_v \leqslant \left(\frac{1-\lambda}{\lambda}\right)^{\frac{1}{2}} (1-\beta)^n .$$

然而, 当 $\mathbb{S}$ 是无限时, 不同于 Doeblin 估计, (5.1.15) 式无法给出一个不依赖于 $\boldsymbol{\mu}$ 的收敛速度. 事实上, 除非 $\sum_i \frac{(\boldsymbol{\mu})_i^2}{(\boldsymbol{\pi})_i} < \infty$, 否则不能得到任何信息. 因此, 有人可能会问当 Doeblin 估计可以发挥作用而且有时做得可能更好时, 为什么还要考虑 (5.1.15) 型的估计. 回答是, 虽然 Doeblin 估计可能很好地发挥作用, 但是当 $\mathbb{S}$ 无限时, 它很少起作用. 即使当 $\mathbb{S}$ 有限时, 通常它也不能给出一个最佳的收敛速度. 请参看后面的习题 5.6.15.

5.2 Dirichlet 型和 β 的估计

本节的目标是寻找使 (5.1.12) 式成立的最佳 (即最大) 值 β 的估计方法. 我们假设 Markov 链是不可约的、可逆的. 后面, 我们还会加上非周期的假设.

5.2.1. Dirichlet 型和 Poincaré不等式: 我们首先需要找到 (5.1.13) 式右边的另一种表示式. 为此, 注意到

$$1-\beta=\sup\{\|\boldsymbol{P}f\|_{2,\boldsymbol{\pi}}:f\in L_0^2(\boldsymbol{\pi}),\|f\|_{2,\boldsymbol{\pi}}=1\}, \tag{5.2.1}$$

其中 $L_0^2(\boldsymbol{\pi})\equiv\{f\in L^2(\boldsymbol{\pi}):\langle f\rangle_{\boldsymbol{\pi}}=0\}$. 显然, (5.1.13) 式的右边的上确界控制了上式右边的上确界. 另一方面, 如果 $f\in L^2(\boldsymbol{\pi})$, $\|f\|_{2,\boldsymbol{\pi}}=1$, 则要么 $\|f-\langle f\rangle_{\boldsymbol{\pi}}\|_{2,\boldsymbol{\pi}}=0$, 从而得到 $\|\boldsymbol{P}f-\langle f\rangle_{\boldsymbol{\pi}}\|_{2,\boldsymbol{\pi}}=0$; 要么 $1\geqslant\|f-\langle f\rangle_{\boldsymbol{\pi}}\|_{2,\boldsymbol{\pi}}>0$, 从而

$$\|\boldsymbol{P}f-\langle f\rangle_{\boldsymbol{\pi}}\|_{2,\boldsymbol{\pi}}\leqslant\left\|\boldsymbol{P}\left(\frac{f-\langle f\rangle_{\boldsymbol{\pi}}}{\|f-\langle f\rangle_{\boldsymbol{\pi}}\|_{2,\boldsymbol{\pi}}}\right)\right\|_{2,\boldsymbol{\pi}}$$

也被 (5.2.1) 式的右边所控制.

进一步, 我们需要 Hilbert 空间上对称算子理论的一个简单结果. 即

$$\begin{aligned}&\sup\{\|\boldsymbol{P}f\|_{2,\boldsymbol{\pi}}:f\in L_0^2(\boldsymbol{\pi}),\|f\|_{2,\boldsymbol{\pi}}=1\}\\=&\sup\{|\langle f,\boldsymbol{P}f\rangle_{\boldsymbol{\pi}}|:f\in L_0^2(\boldsymbol{\pi}),\|f\|_{2,\boldsymbol{\pi}}=1\}.\end{aligned} \tag{5.2.2}$$

由 Schwarz 不等式 $|\langle f,\boldsymbol{P}f\rangle_{\boldsymbol{\pi}}|\leqslant\|f\|_{2,\boldsymbol{\pi}}\|\boldsymbol{P}f\|_{2,\boldsymbol{\pi}}$, (5.2.2) 式的右边被左边所控制. 为了证明相反的不等式, 令 $f\in L_0^2(\boldsymbol{\pi})$, $\|f\|_{2,\boldsymbol{\pi}}=1$. 假设 $\|\boldsymbol{P}f\|_{2,\boldsymbol{\pi}}>0$, 记 $g=\dfrac{\boldsymbol{P}f}{\|\boldsymbol{P}f\|_{2,\boldsymbol{\pi}}}$. 那么 $g\in L_0^2(\boldsymbol{\pi})$ 且 $\|g\|_{2,\boldsymbol{\pi}}=1$. 因此, 如果 γ 表示 (5.2.2) 式右边的上确界, 则由 $\boldsymbol{P}$ 的对称性可知

$$\begin{aligned}4\|\boldsymbol{P}f\|_{2,\boldsymbol{\pi}}&=4\langle g,\boldsymbol{P}f\rangle_{\boldsymbol{\pi}}=\langle(f+g),\boldsymbol{P}(f+g)\rangle_{\boldsymbol{\pi}}-\langle(f-g),\boldsymbol{P}(f-g)\rangle_{\boldsymbol{\pi}}\\&\leqslant\gamma(\|f+g\|_{2,\boldsymbol{\pi}}^2+\|f-g\|_{2,\boldsymbol{\pi}}^2)=2\gamma(\|f\|_{2,\boldsymbol{\pi}}^2+\|g\|_{2,\boldsymbol{\pi}}^2)=4\gamma.\end{aligned}$$

有了 (5.2.2), 再结合 (5.2.1), 我们有

$$\beta=\beta_+\wedge\beta_-, \tag{5.2.3}$$

其中 $\beta_{\pm} \equiv \inf\{\langle f, (\boldsymbol{I} \mp \boldsymbol{P})f\rangle_{\boldsymbol{\pi}} : f \in L_0^2(\boldsymbol{\pi}), \|f\|_{2,\boldsymbol{\pi}} = 1\}$. 由 $(\boldsymbol{I}-\boldsymbol{P})f = (\boldsymbol{I}-\boldsymbol{P})(f - \langle f\rangle_{\boldsymbol{\pi}})$ 可知

$$\beta_{+} = \inf\{\langle f, (\boldsymbol{I} - \boldsymbol{P})f\rangle_{\boldsymbol{\pi}} : f \in L^2(\boldsymbol{\pi}), \mathrm{Var}_{\boldsymbol{\pi}}(f) \equiv \|f - \langle f\rangle_{\boldsymbol{\pi}}\|_{\boldsymbol{\pi}}^2 = 1\}. \tag{5.2.4}$$

同时, 因为对 $f \in L_0^2(\boldsymbol{\pi}), c \in \mathbb{R}$,

$$\langle (f+c), (\boldsymbol{I}+\boldsymbol{P})(f+c)\rangle_{\boldsymbol{\pi}} = \langle f_n, (\boldsymbol{I}+\boldsymbol{P})f_n\rangle_{\boldsymbol{\pi}} + c^2 \ ,$$

所以不论考虑所有的 $f \in L^2(\boldsymbol{\pi})$ 还是所有的 $f \in L_0^2(\boldsymbol{\pi})$, β_- 定义里的下确界显然是一样的. 因此, β_- 的另一个表示是

$$\beta_{-} = \inf\{\langle f, (\boldsymbol{I} + \boldsymbol{P})f\rangle_{\boldsymbol{\pi}} : f \in L^2(\boldsymbol{\pi}), \mathrm{Var}_{\boldsymbol{\pi}}(f) = 1\}. \tag{5.2.5}$$

这里需要做两点说明. 首先, 当 $\boldsymbol{P}$ 为非负定的时(简记为 $\boldsymbol{P} \geqslant 0$), 也即对所有的 $f \in L^2(\boldsymbol{\pi})$ 有 $\langle f, \boldsymbol{P}f\rangle_{\boldsymbol{\pi}} \geqslant 0$ 时, 有 $\beta_+ \leqslant 1 \leqslant \beta_-$, 且因此 $\beta = \beta_+$. 于是

$$\boldsymbol{P} \geqslant 0 \implies \beta = \inf\{\langle f, (\boldsymbol{I} - \boldsymbol{P})f\rangle_{\boldsymbol{\pi}} : f \in L^2(\boldsymbol{\pi}), \mathrm{Var}_{\boldsymbol{\pi}}(f) = 1\}. \tag{5.2.6}$$

其次, 由定理 5.1.14 可知, 除非 Markov 链是非周期的, 都有 $\beta_- = 0$ (参看 (5.1.11) 式和 §5.1.4 开头的叙述), 且当 $\mathbb{S}$ 有限时, $\beta > 0$ 当且仅当 $\mathbb{S}$ 是非周期的.

从 (5.2.4) 和 (5.2.5) 式中的 β_+ 和 β_- 的表达式可以得到很重要的计算工具. 注意到, 由 (5.1.1) 式可知

$$\begin{aligned}\langle f, (\boldsymbol{I} - \boldsymbol{P})f\rangle_{\boldsymbol{\pi}} &= \sum_{(i,j)} f(i)(\boldsymbol{\pi})_i(\boldsymbol{P})_{ij}(f(i) - f(j)) \\ &= \sum_{(i,j)} f(i)(\boldsymbol{\pi})_j(\boldsymbol{P})_{ji}(f(i) - f(j)) \\ &= \sum_{(i,j)} f(j)(\boldsymbol{\pi})_i(\boldsymbol{P})_{ij}(f(j) - f(i)),\end{aligned}$$

因此, 当我们把上式的第二个表达式加到最后的表达式上时, 得到

$$\langle f, (\boldsymbol{I} - \boldsymbol{P})f\rangle_{\boldsymbol{\pi}} = \mathcal{E}(f, f) \equiv \frac{1}{2}\sum_{i \neq j}(\boldsymbol{\pi})_i(\boldsymbol{P})_{ij}(f(j) - f(i))^2. \tag{5.2.7}$$

因为二次型 $\mathcal{E}(f, f)$ 是由 Dirichlet 引入的著名二次型 $\frac{1}{2}\int |\nabla f|^2(x)dx$ 的离散情形类比, 故被称为 *Dirichlet* 型. 推广这个类比, β_+ 可解释为

Poincaré 不等式

$$\beta_+ \mathrm{Var}_{\boldsymbol{\pi}}(f) \leqslant \mathcal{E}(f,f), \qquad f \in L^2(\boldsymbol{\pi}) \tag{5.2.8}$$

中的 *Poincaré* 常数

$$\beta_+ = \inf\{\mathcal{E}(f,f) : f \in L^2(\boldsymbol{\pi}), \mathrm{Var}_{\boldsymbol{\pi}}(f) = 1\}. \tag{5.2.9}$$

类似地, 相应于 (5.2.5) 式, 注意到

$$\begin{aligned}\sum_{(i,j)} (f(j)+f(i))^2(\boldsymbol{\pi})_i(\boldsymbol{P})_{ij} &= \langle \boldsymbol{P}f^2\rangle_{\boldsymbol{\pi}} + 2\langle f, \boldsymbol{P}f\rangle_{\boldsymbol{\pi}} + \|f\|_{2,\boldsymbol{\pi}}^2 \\ &= 2\|f\|_{2,\boldsymbol{\pi}}^2 + 2\langle f, \boldsymbol{P}f\rangle_{\boldsymbol{\pi}}\ ,\end{aligned}$$

因此有

$$\langle f, (\boldsymbol{I}+\boldsymbol{P})f\rangle_{\boldsymbol{\pi}} = \tilde{\mathcal{E}}(f,f) \equiv \frac{1}{2}\sum_{(i,j)} (\boldsymbol{\pi})_i(\boldsymbol{P})_{ij}(f(j)+f(i))^2. \tag{5.2.10}$$

于是

$$\begin{aligned}\beta_- &= \inf\{\tilde{\mathcal{E}}(f,f) : f \in L^2(\boldsymbol{\pi}), \mathrm{Var}_{\boldsymbol{\pi}}(f) = 1\} \\ &= \inf\{\tilde{\mathcal{E}}(f,f) : f \in L_0^2(\boldsymbol{\pi}), \|f\|_{2,\boldsymbol{\pi}} = 1\}.\end{aligned} \tag{5.2.11}$$

为了给出 (5.2.9) 和 (5.2.11) 式的一个直接应用, 假设 (参看习题 2.4.3) 对所有 (i,j), $(\boldsymbol{P})_{ij} \geqslant \epsilon_j$. 令 $\epsilon = \sum\limits_j \epsilon_j$. 假设 $\epsilon > 0$, 并定义一个概率向量 $\boldsymbol{\mu}$ 使得 $(\boldsymbol{\mu})_i = \dfrac{\epsilon_i}{\epsilon}$. 利用 Schwarz 不等式, 同时注意到 $\mathrm{Var}_{\boldsymbol{\pi}}(f)$ 的变分特征值是 $a \rightsquigarrow \langle (f-a)^2\rangle_{\boldsymbol{\pi}}$ 的最小值, 我们有

$$\begin{aligned}2\mathcal{E}(f,f) &= \sum_{(i,j)} (\boldsymbol{\pi})_i(\boldsymbol{P})_{ij}(f(j)-f(i))^2 \\ &\geqslant \epsilon \sum_{(i,j)} (f(j)-f(i))^2(\boldsymbol{\pi})_i(\boldsymbol{\mu})_j \\ &\geqslant \epsilon \sum_i (\langle f\rangle_{\boldsymbol{\mu}} - f(i))^2(\boldsymbol{\pi})_i \geqslant \epsilon \mathrm{Var}_{\boldsymbol{\pi}}(f).\end{aligned}$$

类似地, 有

$$\tilde{\mathcal{E}}(f,f) \geqslant \frac{\epsilon}{2}\mathrm{Var}_{\boldsymbol{\pi}}(f).$$

因此, 由 (5.2.8), (5.2.11) 和 (5.2.3) 式可知, $\beta \geqslant \frac{\epsilon}{2}$. 当然, 这一结论与通过结合习题 2.4.3 和 §5.1.4 中的论证所得的结论相比, 明显要弱得多. 事实上, 由那一习题可知 $\|\delta_i \boldsymbol{P}^n - \boldsymbol{\pi}\|_v \leqslant 2(1-\epsilon)^n$, 结合 §5.1.4 中的论证告诉我们 $\beta \geqslant \epsilon$, 这比我们这里得到的估计好两倍. 另一方面, 正如我们已经指出的那样, 有这样的情形, 我们这里所考察的方法可以得到结果, 而用 Doeblin 型方法得不到.

5.2.2. β_+ 的估计: 许多利用 (5.2.9) 和 (5.2.11) 式去估计 β_+ 和 β_- 的出发点来自下述简单事实:

$$\mathrm{Var}_{\boldsymbol{\pi}}(f) = \frac{1}{2}\sum_{ij}(f(i) - f(j))^2(\boldsymbol{\pi})_i(\boldsymbol{\pi})_j. \tag{5.2.12}$$

这是因为通过展开 $(f(i)-f(j))^2$, 上式右边的和恰好等于 $2\langle f^2\rangle_{\boldsymbol{\pi}} - 2\langle f\rangle_{\boldsymbol{\pi}}^2$.

(5.2.12) 式的重要性在于 $\mathrm{Var}_{\boldsymbol{\pi}}(f)$ 可以通过 $\mathbb{S}$ 上的不同的点上 f 的值之间的差来表示. 显然, $\mathcal{E}(f,f)$ 也是用这样的差表示的. 然而, 出现在 $\mathcal{E}(f,f)$ 中的差仅仅是 f 在每一对使得 $(\boldsymbol{P})_{ij} > 0$ 的 (i,j) 上的值之间的差, 而 (5.2.12) 式的右边包含了所有的抽样点对 (i,j). 因此, 为了通过 $\mathcal{E}(f,f)$ 来估计 $\mathrm{Var}_{\boldsymbol{\pi}}(f)$, 对每一对 (i,j) $(i \neq j)$, 我们需要选择一个**轨道** $p(i,j) = (k_0, \cdots, k_n) \in \mathbb{S}^{n+1}$, 其中 $k_0 = i$, $k_n = j$, 使得对每一 $1 \leqslant m \leqslant n$, $(\boldsymbol{P})_{k_{m-1}k_m} > 0$. 这样的轨道称为**可容许的**. 记

$$(f(i)-f(j))^2 = \left(\sum_{e\in p(i,j)} \triangle_e f\right)^2,$$

其中 $\sum\limits_{e\in p(i,j)}$ 表示 e 取遍轨道 $p(i,j)$ 上的有向线段 (k_{m-1}, k_m), $\triangle_e f \equiv f(\ell) - f(k)$, $e = (k,\ell)$. 有了这一点, 有很多方法可以进行我们的估计. 例如, 给定 $\{\alpha(e) : e \in p(i,j)\} \subseteq (0,\infty)$, 利用 Schwarz 不等式 (参看习题 1.3.1) 可得上式右边不超过

$$\begin{aligned}
&\left(\sum_{e\in p(i,j)} \frac{\alpha(e)}{\rho(e)}\right)\left(\sum_{e\in p(i,j)} \frac{\rho(e)}{\alpha(e)}(\triangle_e f)^2\right) \\
&\leqslant \max_{e\in p(i,j)} \frac{1}{\alpha(e)} \left(\sum_{e\in p(i,j)} \frac{\alpha(e)}{\rho(e)}\right) \sum_{e\in p(i,j)} (\triangle_e f)^2 \rho(e),
\end{aligned}$$

其中, 当 $e = (k,\ell)$ 时 $\rho(e) \equiv (\boldsymbol{\pi})_k(\boldsymbol{P})_{k\ell}$. 因此, 对于轨道集 $\mathcal{P} = \{p(i,j) : (i,j) \in \mathbb{S}^2 \setminus D\}$ (D表示对角线$\{(i,j) \in \mathbb{S}^2 : i = j\}$) 和系数 $\mathcal{A} = \{\alpha(e,p) : e \in p \in \mathcal{P}\} \subseteq (0,\infty)$ 的任意一种选取, 有

$$\begin{aligned}\mathrm{Var}_{\boldsymbol{\pi}}(f) &\leqslant \frac{1}{2}\sum_{p\in\mathcal{P}} w_A(p)\sum_{e\in p}(\triangle_e f)^2\rho(e)\\ &= \frac{1}{2}\sum_{e}(\triangle_e f)^2\rho(e)\left(\sum_{p\ni e} w_A(p)\right) \leqslant W(\mathcal{P},\mathcal{A})\mathcal{E}(f,f),\end{aligned}$$

其中

$$w_A(p) \equiv (\boldsymbol{\pi})_i(\boldsymbol{\pi})_j\left(\max_{e\in p}\frac{1}{\alpha(e,p)}\right)\sum_{e\in p}\frac{\alpha(e,p)}{\rho(e)}, \quad p = p(i,j),$$

和

$$W(\mathcal{P},\mathcal{A}) = \sup_e \sum_{p\ni e} w_{\mathcal{A}}(e,p).$$

因此, 对于可容许轨道集 $\mathcal{P}$ 和系数 $\mathcal{A}$ 的任何一种选取, 我们有

$$\beta_+ \geqslant \frac{1}{W(\mathcal{P},\mathcal{A})}. \tag{5.2.13}$$

(5.2.13) 式的大多数有效应用依赖于 $\mathcal{P}$ 和 $\mathcal{A}$ 的选择, 在一些特殊情形是相当有效的. 给定轨道集的一种取法 $\mathcal{P}$, $\mathcal{A}$ 的一种最常见的选择是取 $\alpha(e,p) \equiv 1$. 在这种情形下, 从 (5.2.13) 式可得

$$\beta_+ \geqslant \frac{1}{W(\mathcal{P})}, \quad \text{其中 } W(\mathcal{P}) \equiv \sup_e \sum_{p\ni e}\sum_{e'\in p}\frac{(\boldsymbol{\pi}(p)) - (\boldsymbol{\pi}(p))_+}{\rho(e')}, \tag{5.2.14}$$

此处如果 p 从 i 开始到 j 结束, 那么 $(\boldsymbol{\pi}(p))_- = (\boldsymbol{\pi})_i$ 和 $(\boldsymbol{\pi}(p))_+ = (\boldsymbol{\pi})_j$.

最后, 应当意识到, 一般而言, (5.2.13) 式没有给出任何信息. 事实上, 虽然不可约性保证了对每一对点至少有一个轨道将它们联系起来, 但是当 $\mathbb{S}$ 无限时, 不能保证我们总能找到 $\mathcal{P}$ 和 $\mathcal{A}$ 使得 $W(\mathcal{P},\mathcal{A}) < \infty$. 此外, 即使当 $\mathbb{S}$ 有限时 (因此对每个选择有 $W(\mathcal{P}, \mathcal{A}) < \infty$), 唯有经深思熟虑的选择才可以使得 (5.2.13) 式产生一个好的估计.

5.2.3. β_- 的估计: 与从 (5.2.9) 式出发对 β_+ 进行估计相比, 从 (5.2.11) 式出发对 β_- 进行估计并不那么直接. 首先, 我们已经知道 (参看定理 5.1.14), 除非 Markov 链是非周期的, 否则 $\beta_- = 0$. 因此, 我们将假设 Markov 链是非周期的. 作为非周期性和不可约性的一

个结果, 我们知道 (参看 (3.1.13) 式) 对每个 $i \in \mathbb{S}$, 总存在一个轨道 $p(i) = (k_0, \cdots, k_{2n+1})$, 它是可容许的 (也即对每个 $1 \leqslant m \leqslant 2n+1$, $(\boldsymbol{P})_{k_{m-1}k_m} > 0$), 闭的 (也即 $k_0 = k_{2n+1}$) 且是从 i 出发的 (即 $k_0 = i$). 注意到我们总是认为这个轨道有奇数步. 这是因为当步数是奇数时, 经初等的运算可得

$$2f(i) = \sum_{m=0}^{2n} (-1)^m (f(k_m) + f(k_{m+1})).$$

因此, 如果 $\tilde{\mathcal{P}}$ 是轨道集的这样一种选择, 使得对于每个 $i \in \mathbb{S}$ 的轨道 $p(i)$, 我们能找到相应的关于系数 $\mathcal{A} = \{\alpha(e,p) : e \in p \in \tilde{\mathcal{P}}\} \subseteq (0, \infty)$ 的一种选择. 因此, 类似于上节中的分析,

$$\begin{aligned} 4\|f\|_{2,\boldsymbol{\pi}}^2 &= \sum_i \Bigg(\sum_{e \in p(i)} (-1)^{m(e)} \tilde{\Delta}_e f \Bigg)^2 (\boldsymbol{\pi})_i \\ &\leqslant \sum_{p \in \tilde{\mathcal{P}}} \boldsymbol{\pi}(p) \left(\sum_{e \in p} \frac{\alpha(e,p)}{\rho(e)} \right) \left(\sum_{e \in p} (\tilde{\Delta}_e f)^2 \frac{\rho(e)}{\alpha(e,p)} \right), \end{aligned}$$

其中, 当 $p = (k_0, \cdots, k_{2n+1})$ 时, $\boldsymbol{\pi}(p) = (\boldsymbol{\pi})_{k_0}$, 并且对 $0 \leqslant m \leqslant 2n$, 当 $e = (k_m, k_{m+1})$ 时, $m(e) = m$, $\widetilde{\Delta}_e f \equiv f(k_m) + f(k_{m+1})$, 因此, 若记

$$\widetilde{w}(p) = \boldsymbol{\pi}(p) \left(\max_{e \in p} \frac{1}{\alpha(e,p)} \right) \sum_{e \in p} \frac{\alpha(e,p)}{\rho(e)},$$

则

$$2\|f\|_{2,\boldsymbol{\pi}}^2 \leqslant \widetilde{W}(\widetilde{\mathcal{P}}, \mathcal{A}) \widetilde{\mathcal{E}}(f,f),$$

其中 $\widetilde{W}(\widetilde{\mathcal{P}}, \mathcal{A}) \equiv \sup_e \sum_{p \ni e} \widetilde{w}(p)$. 因此, 由 (5.2.11) 式, 这就证明了: 对任意一种满足所述条件的轨道 $\widetilde{\mathcal{P}}$ 和系数 $\mathcal{A}$ 的选取, 有

$$\beta_- \geqslant \frac{2}{\widetilde{W}(\widetilde{\mathcal{P}}, \mathcal{A})}.$$

若取 $\alpha(e,p) \equiv 1$, 则简化成

$$\beta_- \geqslant \frac{2}{\widetilde{W}_1(\widetilde{\mathcal{P}})}, \tag{5.2.15}$$

其中 $\widetilde{W}_1(\widetilde{\mathcal{P}}) \equiv \sup_e \sum_{p \ni e} \sum_{e' \in p} \dfrac{\boldsymbol{\pi}(p)}{\rho(e')}$.

应该强调一下，上面关于 β_- 的估计方法是有固有缺陷的，似乎无法获得 $\boldsymbol{P}$ 的谱性质，如非负定性. 当 $\boldsymbol{P}$ 是非负定时，$\beta_- \geqslant 1$. 但是似乎不可能从上述论证中得到这样的结论. 另一方面，由于我们真正感兴趣的是 $\beta = \beta_+ \wedge \beta_-$，而由 (5.2.14) 和 (5.2.15) 给出的估计看上去是可互相比较的，所以 (5.2.15) 式的不充分性造成的损失会比人们所担心的小.

5.3 连续时间可逆 Markov 过程

这里我们将考察上述理论的连续时间参数情形的表现形式. 我们将看到，在连续时间参数情形，这一理论更容易、更加漂亮.

我们在第四章中的记号和理论的基础上进行研究. 特别地，矩阵 $\boldsymbol{Q} = \boldsymbol{R}(\boldsymbol{P} - \boldsymbol{I})$，其中 $\boldsymbol{R}$ 为一个对角矩阵，它的对角线上的元素为速率 $\mathcal{R} = \{R_i : i \in \mathbb{S}\}$，$\boldsymbol{P}$ 为一个转移概率矩阵. 下面我们总假设 $\boldsymbol{Q}$ 为一个不可约的矩阵 (参看 (4.4.2))，也即对每一对 $(i,j) \in \mathbb{S}^2$，有某个 $n \geqslant 0$ 使得 $(\boldsymbol{Q}^n)_{ij} > 0$.

5.3.1. 可逆性准则： 令 $\boldsymbol{Q}$ 为一个给定的矩阵，假设对应的 Markov 过程永远不爆炸 (可参看 §4.3.1). 用 $\{\boldsymbol{P}(t) : t \geqslant 0\}$ 表示由 $\boldsymbol{Q}$ 决定的半群 (参看推论 4.3.2). 在这一小节中，我们的目的是证明若 $\hat{\boldsymbol{\mu}}$ 为一个概率向量，使得逐一平衡条件

$$(\hat{\boldsymbol{\mu}})_i(\boldsymbol{Q})_{ij} = (\hat{\boldsymbol{\mu}})_j(\boldsymbol{Q})_{ji}, \quad (i,j) \in \mathbb{S}^2 \tag{5.3.1}$$

成立,则逐一平衡条件

$$(\hat{\boldsymbol{\mu}})_i(\boldsymbol{P}(t))_{ij} = (\hat{\boldsymbol{\mu}})_j(\boldsymbol{P}(t))_{ji}, \quad t > 0, (i,j) \in \mathbb{S}^2 \tag{5.3.2}$$

也成立.

当速率 $\mathcal{R}$ 有界时，从 (5.3.1) 式推出 (5.3.2) 式的证明是平凡的. 事实上，我们只要首先利用一个简单的归纳步骤来验证 (5.3.1) 式推出对一切 $n \geqslant 0$，$(\hat{\boldsymbol{\mu}})_i(\boldsymbol{Q}^n)_{ij} = (\hat{\boldsymbol{\mu}})_j(\boldsymbol{Q}^n)_{ji}$，然后再利用 (4.2.13) 中给出的关于 $\boldsymbol{P}(t)$ 的表达式. 当速率 $\mathcal{R}$ 无界时，我们可用 §4.3.1 中的逼近方法. 确切地，参看 §4.3.1，令 $\boldsymbol{Q}^{(N)}$ 是相应于那里选取的 $\mathcal{R}^{(N)}$ 的矩阵. 也就是说，若 $i \in F_N$，令 $(\boldsymbol{Q}^{(N)})_{ij} = (\boldsymbol{Q})_{ij}$；若 $i \notin F_N$，令 $(\boldsymbol{Q}^{(N)})_{ij} = 0$. 利用归纳法，首先，对一切 $n \geqslant 1$ 和 $i \notin F_N$，有 $((\boldsymbol{Q}^{(N)})^n)_{ij} = 0$；其次，

对一切 $n \geqslant 0$ 和 $(i,j) \in F_N^2$,

$$(\hat{\boldsymbol{\mu}})_i((\boldsymbol{Q}^{(N)})_{ij}^n = (\hat{\boldsymbol{\mu}})_j((\boldsymbol{Q}^{(N)})^n)_{ji} .$$

因此, 若 $\{\boldsymbol{P}^N(t) : t \geqslant 0\}$ 为一个由 $\boldsymbol{Q}^{(N)}$ 决定的半群, 那么因为关于 $\boldsymbol{Q}^{(N)}$ 的速率是有界的, 由 (4.2.13) 式, 得

$$\begin{aligned} &(\boldsymbol{P}^{(N)}(t))_{ij} = \delta_{i,j}, && \text{若 } i \notin F_N, \\ &(\hat{\boldsymbol{\mu}})_i(\boldsymbol{P}^{(N)}(t))_{ij} = (\hat{\boldsymbol{\mu}})_j(\boldsymbol{P}^{(N)}(t))_{ji}, && \text{若 } (i,j) \in F_N^2. \end{aligned} \tag{5.3.3}$$

特别地, 因为根据 (4.3.3) 式, 有

$$(\boldsymbol{P}^{(N)}(t))_{ij} \to (\boldsymbol{P}(t))_{ij},$$

所以结论得证.

作为上述推导的一个推论, 由 (5.3.1) 式可推出 $\hat{\boldsymbol{\mu}}$ 关于 $\boldsymbol{P}(t)$ 是平稳的. 因此, 由于我们假设 $\boldsymbol{Q}$ 为一个不可约的矩阵, §4.4.2 和 §4.4.3 中的结果保证了我们可以把 $\hat{\boldsymbol{\mu}}$ 作为定理 4.4.8 中引入并在 §4.4.3 中得到讨论 (参看 (4.4.11)) 的概率向量 $\hat{\boldsymbol{\pi}} \equiv \hat{\boldsymbol{\pi}}^{\mathbb{S}}$. 综合上述, 若 $\hat{\boldsymbol{\mu}}$ 是一个使得 (5.3.1) 成立的概率向量, 则 $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\pi}}$.

5.3.2. 有界速率时 $L^2(\hat{\boldsymbol{\pi}})$ 中的收敛性: 由刚才得到的结论,我们现在假设 $\hat{\boldsymbol{\pi}}$ 是一个概率向量, 使得当 $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\pi}}$ 时 (5.3.1) 式成立. 特别地, 这意味着对一切 $t > 0$, 有

$$(\hat{\boldsymbol{\pi}})_i(\boldsymbol{P}(t))_{ij} = (\hat{\boldsymbol{\pi}})_j(\boldsymbol{P}(t))_{ji}, \quad (i,j) \in \mathbb{S}^2. \tag{5.3.4}$$

在有了 (5.3.4) 式后, 我们可以利用 §5.2 中的思想去得到作为 $L^2(\hat{\boldsymbol{\pi}})$ 中的收敛性之度量的 $\boldsymbol{P}(t)f \to \langle f\rangle_{\hat{\boldsymbol{\pi}}}$ 的收敛速度的一个估计. 为了详细讨论, 首先注意到, 对任意的 $h > 0$,

$$\langle f, \boldsymbol{P}(h)f\rangle_{\hat{\boldsymbol{\pi}}} = \left\|\boldsymbol{P}\left(\frac{h}{2}\right) f\right\|_{2,\hat{\boldsymbol{\pi}}}^2 \geqslant 0,$$

因此 (参看 (5.1.13), (5.2.6) 和 (5.2.7))

$$\begin{aligned} \beta(h) &\equiv 1 - \sup\{|\langle f, \boldsymbol{P}(h)f\rangle_{\hat{\boldsymbol{\pi}}}| : f \in L_0^2(\hat{\boldsymbol{\pi}}),\ \|f\|_{2,\hat{\boldsymbol{\pi}}} = 1\} \\ &= \inf\{\langle f, (\boldsymbol{I} - \boldsymbol{P}(h))f\rangle_{\hat{\boldsymbol{\pi}}} : \mathrm{Var}_{\hat{\boldsymbol{\pi}}}(f) = 1\} \\ &= \inf\{\mathcal{E}_h(f,f) : \mathrm{Var}_{\hat{\boldsymbol{\pi}}}(f) = 1\}, \end{aligned}$$

其中

$$\mathcal{E}_h(f,f) \equiv \frac{1}{2}\sum_{i\neq j}(\hat{\boldsymbol{\pi}})_i(\boldsymbol{P}(h))_{ij}(f(j)-f(i))^2$$

是 $L^2(\hat{\boldsymbol{\pi}})$ 上 $\boldsymbol{P}(h)$ 的 Dirichlet 型. 因此 (参看 (5.1.12)), 对任意的 $t>0$ 和 $n\in\mathbb{Z}^+$, 有

$$\|\boldsymbol{P}(h)f-\langle f\rangle_{\hat{\boldsymbol{\pi}}}\|_{L^2(\hat{\boldsymbol{\pi}})} \leqslant \left(1-\beta\left(\frac{t}{n}\right)\right)^n \|f-\langle f\rangle_{\hat{\boldsymbol{\pi}}}\|_{2,\hat{\boldsymbol{\pi}}}.$$

下面我们附加假设: 其速率 $\mathcal{R}$ 是有界的. 这时, 由 (4.2.8) 式可以看出: 对满足 $\mathrm{Var}_{\hat{\boldsymbol{\pi}}}(f)=1$ 的 f, 一致地有

$$\lim_{h\searrow 0}\frac{\mathcal{E}_h(f,f)}{h}=\mathcal{E}^{\boldsymbol{Q}}(f,f),$$

其中

$$\mathcal{E}^{\boldsymbol{Q}}(f,f) \equiv \frac{1}{2}\sum_{i\neq j}(\hat{\boldsymbol{\pi}})_i(\boldsymbol{Q})_{ij}(f(j)-f(i))^2; \tag{5.3.5}$$

由此又有极限 $\lim\limits_{h\searrow 0}\dfrac{\beta(h)}{h}$ 存在并且等于

$$\lambda \equiv \inf\{\mathcal{E}^{\boldsymbol{Q}}(f,f): f\in L^2(\boldsymbol{\pi}), \mathrm{Var}_{\hat{\boldsymbol{\pi}}}(f)=1\}. \tag{5.3.6}$$

因此至少当速率有界时, 我们有

$$\|\boldsymbol{P}(h)f-\langle f\rangle_{\hat{\boldsymbol{\pi}}}\|_{2,\hat{\boldsymbol{\pi}}} \leqslant e^{-\lambda t}\|f-\langle f\rangle_{\hat{\boldsymbol{\pi}}}\|_{2,\hat{\boldsymbol{\pi}}}. \tag{5.3.7}$$

5.3.3. 一般情形下 $L^2(\hat{\boldsymbol{\pi}})$-收敛速度: 当速率 $\mathcal{R}$ 无界时, 上面的推理过程是缺乏根据的. 为了处理无界情形, 我们需要某些附加的讨论. 这种讨论源于下面的引理[3].

5.3.8 引理. 给定 $f\in L^2(\hat{\boldsymbol{\pi}})$, 函数 $t\in[0,\infty)\mapsto\|\boldsymbol{P}(h)f\|^2_{2,\hat{\boldsymbol{\pi}}}$ 是连续、非增、非负并且凸的. 特别地, 函数 $t\in(0,\infty)\mapsto\dfrac{\langle f,(\boldsymbol{I}-\boldsymbol{P}(h))f\rangle_{\hat{\boldsymbol{\pi}}}}{t}$ 是非增的, 从而

$$\lim_{h\searrow 0}\frac{\|f\|^2_{2,\hat{\boldsymbol{\pi}}}-\|\boldsymbol{P}(h)f\|^2_{2,\hat{\boldsymbol{\pi}}}}{h}$$

在 $[0,\infty]$ 上存在. 事实上, 我们有 (参看 (5.3.5))

$$\lim_{h\searrow 0}\frac{\|f\|^2_{2,\hat{\boldsymbol{\pi}}}-\|\boldsymbol{P}(h)f\|^2_{2,\hat{\boldsymbol{\pi}}}}{h}\geqslant 2\mathcal{E}^{\boldsymbol{Q}}(f,f).$$

---

[3] 如果人们熟悉谱理论, 特别是 Stone 定理, 那么相当麻烦的论证是可以被避免的.

证明: 对给定的 $f \in L^2(\hat{\pi})$, 令 $f_N = \mathbf{1}_{F_N} f$ (参看 §4.3.1 中的记号). 由 Lebesgue 控制收敛定理得, 当 $N \to \infty$ 时, $\|f - f_N\|_{2,\hat{\pi}} \to 0$. 因为对任意的 $t > 0$, $g \in L^2(\hat{\pi})$, 有 $\|\boldsymbol{P}(t)g\|_{2,\hat{\pi}} \leqslant \|g\|_{2,\hat{\pi}}$, 所以当 $N \to \infty$ 时, 在 $t \in (0,\infty)$ 上一致地有

$$\left| \|\boldsymbol{P}(t)f\|_{2,\hat{\pi}} - \|\boldsymbol{P}(t)f_N\|_{2,\hat{\pi}} \right| \leqslant \|\boldsymbol{P}(t)(f - f_N)\|_{2,\hat{\pi}} \leqslant \|f - f_N\|_{2,\hat{\pi}} \to 0.$$

因此由后面的习题 5.6.1(a) 可知, 为了证明最初的论断, 我们只需证明对某个 M, f 在 F_M 以外为零时该论断成立. 下面就设函数 f 在 F_M 之外为零, 记 $\psi(t) = \|\boldsymbol{P}(t)f\|_{2,\hat{\pi}}^2$, 又记当 $N \geqslant M$ 时 $\psi_N(t) = \|\boldsymbol{P}^N(t)f\|_{2,\hat{\pi}}^2$. 由 (4.3.3) 可知, 在有限区间上 $\psi_N \to \psi$ 一致成立. 再次应用习题 5.6.1 (a) 可知, 代替 ψ, 我们只需考虑 ψ_N 就够了. 也就是说, 只要我们证明 ψ_N 是一个非增、非负并且凸的函数, 那么, ψ 也是这样的一个函数. 非负性是显然的. 为证明其他性质, 利用 (4.2.7) 式得

$$\dot{\psi}_N(t) = \langle \boldsymbol{Q}^{(N)}\boldsymbol{P}^{(N)}(t)f, \boldsymbol{P}^{(N)}(t)f\rangle_{\hat{\pi}} + \langle \boldsymbol{P}^{(N)}(t)f, \boldsymbol{Q}^{(N)}\boldsymbol{P}^{(N)}(t)f\rangle_{\hat{\pi}}.$$

其次, 由 (5.3.3) 的第一行可知, 因为 $N \geqslant M$, $\boldsymbol{P}^N(t)f$ 在 F_M 以外等于零. 因此由于 $(\hat{\pi})_i \boldsymbol{Q}_{ij}^N = (\hat{\pi})_j \boldsymbol{Q}_{ji}^{(N)}$, $(i,j) \in F_N^2$, 我们得

$$\dot{\psi}_N(t) = 2\langle \boldsymbol{P}^{(N)}(t)f, \boldsymbol{Q}^{(N)}\boldsymbol{P}^{(N)}(t)f\rangle_{\hat{\pi}}.$$

同理可得

$$\begin{aligned}\ddot{\psi}_N(t) &= 2\langle \boldsymbol{Q}^{(N)}\boldsymbol{P}^{(N)}(t)f, \boldsymbol{Q}^{(N)}\boldsymbol{P}^{(N)}(t)f\rangle_{\hat{\pi}} + \\ &\quad\ 2\langle \boldsymbol{P}^{(N)}(t)f, (\boldsymbol{Q}^{(N)})^2\boldsymbol{P}^{(N)}(t)f\rangle_{\hat{\pi}} \\ &= 4\langle \boldsymbol{Q}^{(N)}\boldsymbol{P}^{(N)}(t)f, \boldsymbol{Q}^{(N)}\boldsymbol{P}^{(N)}(t)f\rangle_{\hat{\pi}} \\ &= 4\|\boldsymbol{Q}^{(N)}\boldsymbol{P}^{(N)}(t)f\|_{2,\hat{\pi}}^2 \geqslant 0.\end{aligned}$$

很明显, 其中第二式表明 ψ_N 的凸性. 为了证明第一式推出 ψ_N 是非增的. 我们将证明

$$\begin{aligned}\langle g, -\boldsymbol{Q}^{(N)}g\rangle_{\hat{\pi}} = \frac{1}{2} \sum_{(i,j)\in F_N^2, i\neq j} (\hat{\pi})_i(\boldsymbol{Q})_{ij}(g(j) - g(i))^2 + \\ \sum_{i\in F_N} (\hat{\pi})_i V^{(N)}(i) g(i)^2 \end{aligned} \tag{$*$}$$

其中对 $i \in F_N$, $V^{(N)}(i) \equiv \sum_{j \notin F_N} \boldsymbol{Q}_{ij}$; 而对 $i \notin F_N$, $g(i) = 0$. 为了验证 $(*)$ 式, 首先注意到,

$$\begin{aligned}\langle g, -\boldsymbol{Q}^{(N)} g\rangle_{\hat{\boldsymbol{\pi}}} &= -\sum_{(i,j)\in F_N^2} (\hat{\boldsymbol{\pi}})_i(\boldsymbol{Q})_{ij} g(i)g(j)\\ &= -\sum_{(i,j)\in F_N^2, i\neq j} (\hat{\boldsymbol{\pi}})_i(\boldsymbol{Q})_{ij} g(i)(g(j)-g(i)) +\\ &\qquad \sum_{i\in F_N} (\hat{\boldsymbol{\pi}})_i V^{(N)}(i) g(i)^2.\end{aligned}$$

利用 $(\hat{\boldsymbol{\pi}})_i \boldsymbol{Q}_{ij} = (\hat{\boldsymbol{\pi}})_j \boldsymbol{Q}_{ji}$, $(i,j) \in F_N^2$, 我们有

$$\begin{aligned}&-\sum_{(i,j)\in F_N^2, i\neq j} (\hat{\boldsymbol{\pi}})_i(\boldsymbol{Q})_{ij} g(i)(g(j)-g(i))\\ =& \sum_{(i,j)\in F_N^2, i\neq j} (\hat{\boldsymbol{\pi}})_i(\boldsymbol{Q})_{ij} g(i)(g(j)-g(i)).\end{aligned}$$

因此 $(*)$ 式得证. 最后, 在 $(*)$ 式中取 $g = \boldsymbol{P}^N(t)f$ 即得 $\dot{\psi}_N(t) \leqslant 0$.

下面转到第二和第三个结论. 令 f 是 $L^2(\hat{\pi})$ 中的任意一个元. 我们知道相应的 ψ 是一个非增、连续、非负并且凸的函数. 容易验证 (参看习题 5.6.1(**d**)), $t \rightsquigarrow \dfrac{\psi(0)-\psi(t)}{t}$ 是非增的, 因此

$$\lim_{h\searrow 0} \frac{\psi(0)-\psi(h)}{h}$$

在 $[0,\infty]$ 上存在. 其次, 回顾 (5.2.7) 式, 取其中的 $\boldsymbol{P} = \boldsymbol{P}(2h)$, 则有

$$\psi(0)-\psi(h) = \langle f, (\boldsymbol{I}-\boldsymbol{P}(2h))f\rangle_{\hat{\boldsymbol{\pi}}} = \frac{1}{2}\sum_{(i,j), i\neq j} (\hat{\boldsymbol{\pi}})_i(\boldsymbol{P}(2h))_{ij}(f(j)-f(i))^2.$$

因此, 并由 (4.3.4) 式有

$$\lim_{h\searrow 0} \frac{(\boldsymbol{P}(2h))_{ij}}{h} = 2(\boldsymbol{Q})_{ij}, \quad i \neq j,$$

再应用 Fatou 引理, 即得欲证的不等式. □

5.3.9 引理. 若 $0 < s < t$, 则对任意的 $f \in L^2(\hat{\boldsymbol{\pi}})$,

$$\frac{\|f\|_{2,\hat{\boldsymbol{\pi}}}^2}{s} \geqslant \frac{\|\boldsymbol{P}(s)f\|_{2,\hat{\boldsymbol{\pi}}}^2 - \|\boldsymbol{P}(t)f\|_{2,\hat{\boldsymbol{\pi}}}^2}{t-s} \geqslant 2\mathcal{E}^{\boldsymbol{Q}}(\boldsymbol{P}(t)f, \boldsymbol{P}(t)f).$$

证明: 记 $\psi(t)=\|\boldsymbol{P}(t)f\|_{2,\hat{\pi}}^2$. 则 ψ 为一个非增、连续、非负并且凸的函数. 因此由习题 5.6.1 (a), 对任意 $h>0$, 得

$$\frac{\psi(0)}{s}\geqslant\frac{\psi(0)-\psi(s)}{s}\geqslant\frac{\psi(s)-\psi(t)}{s-t}\geqslant\frac{\psi(t)-\psi(t+h)}{h}.$$

此外, 由于

$$\frac{\psi(t)-\psi(t+h)}{h}=\frac{\|\boldsymbol{P}(t)f\|_{2,\hat{\pi}}^2-\|\boldsymbol{P}(h)\boldsymbol{P}(t)f\|_{2,\hat{\pi}}^2}{h},$$

再将引理 5.3.8 的最后一部分中的 f 用 $\boldsymbol{P}(t)f$ 代替, 即得引理的第二个结论. □

有了上面的结论, 我们就可以来完成本节的计划了.

$$\begin{aligned}&\|f\|_{2,\hat{\pi}}^2-\|\boldsymbol{P}(t)f\|_{2,\hat{\pi}}^2\\&=\sum_{m=0}^{n-1}\left(\left\|\boldsymbol{P}\left(\frac{mt}{n}\right)f\right\|_{2,\hat{\pi}}^2-\left\|\boldsymbol{P}\left(\frac{(m+1)t}{n}\right)f\right\|_{2,\hat{\pi}}^2\right),\end{aligned}$$

利用引理 5.3.9 的结果可得估计

$$\|f\|_{2,\hat{\pi}}^2-\|\boldsymbol{P}(t)f\|_{2,\hat{\pi}}^2\geqslant\frac{2t}{n}\sum_{m=1}^{n}\mathcal{E}^{\boldsymbol{Q}}\left(\boldsymbol{P}\left(\frac{mt}{n}\right)f,\boldsymbol{P}\left(\frac{mt}{n}\right)f\right).$$

因此对任意的 $f\in L_0^2(\hat{\pi})$,

$$\|f\|_{2,\hat{\pi}}^2-\|\boldsymbol{P}(t)f\|_{2,\hat{\pi}}^2\geqslant\frac{2\lambda t}{n}\sum_{m=1}^{n}\left\|\boldsymbol{P}\left(\frac{mt}{n}\right)f\right\|_{2,\hat{\pi}}^2,$$

其中 λ 的定义参看 (5.3.6) 式. 令 $n\to\infty$, 得到

$$\|f\|_{2,\hat{\pi}}^2-\|\boldsymbol{P}(t)f\|_{2,\hat{\pi}}^2\geqslant 2\lambda\int_0^t\|\boldsymbol{P}(\tau)f\|_{2,\hat{\pi}}^2d\tau.$$

最后, 利用 *Gronwall* 不等式 (参看习题 5.6.4), 由上面的论证得到估计 $\|\boldsymbol{P}(t)f\|_{2,\hat{\pi}}^2\leqslant e^{-2\lambda t}\|f\|_{2,\hat{\pi}}^2$. 用 $f-\langle f\rangle_{\hat{\pi}}$ 代替 $f\in L^2(\hat{\pi})$ 就得 (5.3.7) 式在 $\mathcal{R}$ 无界时也成立.

5.3.4. 估计 λ: 利用与 §5.2.2 中估计 β_+ 时完全相同的方式我们可以来估计 (5.3.6) 式中的 λ. 也就是说, 我们选取一个由轨道 $p(i,j)$ 组成的 $\mathcal{P}$, 使得对于每对 $(i,j)\in\mathbb{S}\backslash D$, 具有以下性质: 如果 $p(i,j)=(k_0,\cdots,k_n)$,

则有 $k_0 = i, k_n = j$, 且 $p(i,j)$ 是可容许的, 也即对任一 $1 \leqslant m \leqslant n$, $(\boldsymbol{Q})_{k_{m-1}k_m} > 0$. 于是, 如同在 §5.2.2 中证明的, 我们有

$$\lambda \geqslant \frac{1}{W(\mathcal{P})}, \tag{5.3.10}$$

其中

$$W(\mathcal{P}) \equiv \sup_{e} \sum_{p \ni e} \sum_{e' \in p} \frac{(\hat{\boldsymbol{\pi}}(p))_{-}(\hat{\boldsymbol{\pi}}(p))_{+}}{\rho(e')},$$

这里的上确界是取遍所有使得 $(\boldsymbol{Q})_{k\ell} > 0$ 的有向边 $e = (k, \ell)$; 第一个求和是对含有边 e 的轨道 $p \in \mathcal{P}$ 进行的; 第二个求和是对出现在轨道 p 中的边 e' 进行的; 若 p 从 i 开始, 则 $(\hat{\boldsymbol{\pi}}(p))_{-} = (\hat{\boldsymbol{\pi}})_i$; 若 p 在 j 结束, 则 $(\hat{\boldsymbol{\pi}}(p))_{+} = (\hat{\boldsymbol{\pi}})_j$; 若 $e' = (k, \ell)$, 则 $\rho(e') = (\hat{\boldsymbol{\pi}})_k(\boldsymbol{Q})_{k\ell}$.

5.4 Gibbs 态和 Glauber 动力系统

粗略地说, 统计力学的物理原理可以概括如下: 当一个系统处于平衡时, *低能状态比高能状态更加可靠*. 事实上, Gibbs 精确化了上述提法: 状态 i 的概率与 $e^{-\frac{H(i)}{kT}}$ 成正比, 其中 k 是 Boltzmann 常数, T 是温度, 而 $H(i)$ 是系统在状态 i 时的能量. 由这种 Gibbs 方式赋予概率的分布称为 *Gibbs 态*.

由于 Gibbs 态是针对于平衡的模型, 有理由问动力系统平衡的条件是什么. 从我们的观点看, 这意味着要寻找一个 Markov 过程使得 Gibbs 态为其平稳分布. 进一步, 由于动力系统在物理学意义上是可逆的, 我们要寻找 Markov 过程使得相对于 Gibbs 态它是可逆的, 并且, 因为上述这类过程是由 Glauber 引入的, 因此我们称相对于 Gibbs 态是可逆的 Markov 过程为关于这个 Gibbs 态的 *Glauber 动力系统*.

在这一节里, 我们将简述 Gibbs 态以及与之相关的 Glauber 动力系统.

5.4.1. 框架: 在这一节中, 我们将使用如下记号. 记 $\mathbb{S}$ 为一个有限或可数无限的空间. 在 $\mathbb{S}$ 上存在一个在某种 “自然” 的背景下给定的 “权” $\boldsymbol{\nu} \in (0, \infty)^{\mathbb{S}}$, 它可以看做一个行向量 (不一定可和). 在许多应用场合中 $\boldsymbol{\nu}$ 常常是均匀的, 即它的每一个分量都是 1, 但在另一些场合, 不假定它是均匀的更为方便. 其次, 设有一个函数 $H : \mathbb{S} \to [0, \infty)$ (又称能量

函数), 它有以下性质: 对任一 $\beta \in (0, \infty)$,

$$Z(\beta) \equiv \sum_{i\in\mathbb{S}} e^{-\beta H(i)}(\boldsymbol{\nu})_i < \infty. \tag{5.4.1}$$

按物理学的隐喻, 如不考虑 Boltzmann 常数 k, $\beta = \frac{1}{kT}$ 可看做是温度倒数. 物理学家称 $\beta \rightsquigarrow Z(\beta)$ 为*划分函数*. 最后, 对于任一 $\beta \in (0, \infty)$, *Gibbs 态* $\gamma(\beta)$ 为由下式给定的一个概率向量

$$(\gamma(\beta))_i = \frac{1}{Z(\beta)} e^{-\beta H(i)}(\boldsymbol{\nu})_i, \qquad i \in \mathbb{S}. \tag{5.4.2}$$

从物理学的观点看, 每一个有兴趣的量都可用划分函数来表示. 例如, 最基本的是通过取对数求导计算能量的均值和方差:

$$\langle H\rangle_{\gamma(\beta)} = -\frac{d}{d\beta}\log Z(\beta), \quad \mathrm{Var}_{\gamma(\beta)}(H) = \frac{d^2}{d\beta^2}\log Z(\beta). \tag{5.4.3}$$

最后, 我们来讨论 Glauber 动力系统的表述. 为此, 我们从一个矩阵 $\boldsymbol{A}$ 开始, 它的元素均为非负且对角元素都是 0. 进一步假设 $\boldsymbol{A}$ 是不可约的, 也即对任意的 $(i, j) \in \mathbb{S}^2$,

$$\sup_{n\geqslant 0}(\boldsymbol{A}^n)_{ij} > 0, \tag{5.4.4}$$

而且它是可逆的, 即对任意的 $(i, j) \in \mathbb{S}^2$,

$$(\boldsymbol{\nu})_i(\boldsymbol{A})_{ij} = (\boldsymbol{\nu})_j(\boldsymbol{A})_{ji} . \tag{5.4.5}$$

最后, 假设对任意的 $i \in \mathbb{S}, \beta > 0$, 有

$$\sum_{j\in\mathbb{S}} e^{-\beta H(j)}(\boldsymbol{A})_{ij} < \infty. \tag{5.4.6}$$

有了这些, 我们有许多方法去构造一个 Glauber 动力系统. 然而, 就我们的目的而言, 最适合的是具有如下 Q-矩阵的那一个:

$$\begin{aligned}(\boldsymbol{Q}(\beta))_{ij} &= e^{-\beta(H(j)-H(i))^+}(\boldsymbol{A})_{ij}, \qquad \text{当 } j \neq i,\\ (\boldsymbol{Q}(\beta))_{ii} &= -\sum_{j\neq i}(\boldsymbol{Q}(\beta))_{ij},\end{aligned} \tag{5.4.7}$$

其中 $a^+ \equiv a \vee 0$ 为实数 $a \in \mathbb{R}$ 的非负部分. 因为对 $i \neq j$,

$$(\gamma(\beta))_i(\boldsymbol{Q}(\beta))_{ij} = Z(\beta)^{-1}e^{-\beta(H(i)\vee H(j))}(\boldsymbol{\nu})_i(\boldsymbol{A})_{ij}, \tag{5.4.8}$$

所以显然 $\gamma(\beta)$ 关于 $\boldsymbol{Q}(\beta)$ 可逆. 存在着许多其他的概率向量, 而最优的选择常常是由所考虑的问题的特殊性质决定的. 然而，不管选取什么, 应该使得对于任意一个 $\beta > 0$, $\boldsymbol{Q}(\beta)$ 决定一个从不爆炸的 *Markov* 过程.

5.4.2. Dirichlet 型: 在这一小节中, 我们将修改 §5.2.3 中提出的思想以获得下界:

$$\lambda_\beta \equiv \inf\{\mathcal{E}_\beta(f,f) : \mathrm{Var}_\beta(f) = 1\}, \tag{5.4.9}$$

其中 $\mathcal{E}_\beta(f,f) \equiv \frac{1}{2}\sum\limits_{j\neq i}(\gamma(\beta))_i(\boldsymbol{Q}(\beta))_{ij}(f(j)-f(i))^2$, 这里 $\mathrm{Var}_\beta(f)$ 是 f 关于 $\gamma(\beta)$ 的方差 $\mathrm{Var}_{\gamma(\beta)}(f)$ 的简写. 为此, 引入下面记号

$$\mathrm{Elev}(p) = \max_{0\leqslant m\leqslant n} H(i_m), \quad e(p) = \mathrm{Elev}(p) - H(i_0) - H(i_n),$$

其中 $p = (i_0, \cdots, i_n)$ 是一个轨道. 当 $\boldsymbol{Q}(\beta)$ 由 (5.4.7) 式给出时, 对 $p = (i_0, \cdots, i_n)$, 我们有

$$w_\beta(p) \equiv \sum_{m=1}^{n} \frac{(\gamma(\beta))_{i_0}(\gamma(\beta))_{i_n}}{(\gamma(\beta))_{i_{m-1}}(\boldsymbol{Q}(\beta))_{i_{m-1}i_m}} \leqslant Z(\beta)^{-1}e^{\beta e(p)}w(p),$$

其中

$$w(p) \equiv \sum_{m=1}^{n} \frac{(\boldsymbol{\nu})_{i_0}(\boldsymbol{\nu})_{i_n}}{(\boldsymbol{\nu})_{i_{m-1}}\boldsymbol{A}_{i_{m-1}i_m}}.$$

因此, 对于轨道集 $\mathcal{P}$ 的任一种选取, 我们有 (参看 (5.3.10)),

$$W_\beta(\mathcal{P}) \equiv \sup_e \sum_{p\ni e} w_\beta(p) \leqslant Z(\beta)^{-1}e^{\beta E(\mathcal{P})}W(\mathcal{P}),$$

其中 $W(\mathcal{P}) \equiv \sup\limits_e \sum\limits_{p\ni e} w_\beta(p)$, $E(\mathcal{P}) \equiv \sum\limits_{p\in\mathcal{P}} e(p)$. 从而得到

$$\lambda_\beta \geqslant \frac{Z(\beta)e^{-\beta E(\mathcal{P})}}{W(\mathcal{P})}. \tag{5.4.10}$$

一方面, 显然仅当 $W(\mathcal{P}) < \infty$ 时, (5.4.10) 式才能提供信息. 同时, 它说明: 选取 $\mathcal{R}$ 使 $E(\mathcal{P})$ 尽可能小很重要, 至少当我们感兴趣的是大的 β 的情形是这样. 当 $\mathbb{S}$ 有限时, 协调这两者没有任何问题. 事实上, $\mathbb{S}$ 的有限性保证了对任意选取的可容许轨道集 $W(\mathcal{P})$ 有限. 另外, 有限性保证了对每一 (i,j) 可以找到一条轨道 $p(i,j)$, 使得 $\mathrm{Elev}(p)$ 对所有从 i 到 j 的可容许轨道达到最小, 并且, 显然任一含有所有这

样轨道的 $\mathcal{P}$ 将使得 $E(\mathcal{P})$ 达到最小. 当然, 为使 $W(\mathcal{P})$ 也达到最小需要精心挑选 $\mathcal{P}$. 无论如何, 只要 $\mathbb{S}$ 有限并且 $\mathcal{P}$ 含有使得 $\text{Elev}(p)$ 对所有从 i 到 j 的轨道 p 达到最小的所有轨道 $p(i,j)$, $E(\mathcal{P})$ 就有好的解释. 也即, 把 $\mathbb{S}$ 想象为一张地图中的位置集合, 而 H 为这些位置的海拔高度集合. 就是说, 按这种比喻, $H(i)$ 是 i 高出海平面的距离. 不失一般性, 我们假设至少有一个位置 k_0 在海平面上: $H(k_0)=0$.[4] 当这样的 k_0 存在时, $E(\mathcal{P})$ 的解释为: 不管一个旅行者从何处出发, 途经何种可容许轨道, 为接触到海洋, $E(\mathcal{P})$ 是他必须到达的海拔高度的最小上界. 为了说明这一点, 如果 p 和 p' 是一对可容许轨道, 且 p 的终点是 p' 的起点, 当 q 是 p 和 p' 的连接 (即若 $p=(i_0,\cdots,i_n)$ 和 $p'=(i'_0,\cdots,i'_n)$, 则 $q=(i_0,\cdots,i_n,i'_1,\cdots,i'_n)$) 时, 那么 q 是可容许轨道且 $\text{Elev}(q)\leqslant \text{Elev}(p)\vee\text{Elev}(p')$. 因而, 对任一 (i,j), $e(p(i,j))=e(p(i,k_0))\vee e(p(j,k_0))$, 由此易得 $E(\mathcal{P})=\max\limits_i e(p(i,k_0))$. 最后, 由于对每一 i, $e(p(i,k_0))=H(\ell)-H(i)$, 其中 ℓ 是轨道 $p(i,k_0)$ 上的最高点, 这就完成了我们的说明. 当 $\mathbb{S}$ 无限时, 在许多场合同样的解释仍然有效. 例如, 这样的解释适应于 H 在无穷远处趋于无穷的情形, 也就是, 对任意的 $M<\infty$, $\{i: H(i)\leqslant M\}$ 有限的情形.

当 $\mathbb{S}$ 有限时, 我们可以证明, 至少对充分大的 β, (5.4.10) 式是相当不错的. 更精确地, 我们有如下的结果.

5.4.11 定理. *假设 $\mathbb{S}$ 是有限的, $\boldsymbol{Q}(\beta)$ 由 (5.4.7) 式定义. 令 $\boldsymbol{m}=\min\limits_{i\in\mathbb{S}}H(i)$, $\mathbb{S}_0=\{i:H(i)=\boldsymbol{m}\}$, $\boldsymbol{e}$ 为当 $\mathcal{P}$ 取遍所有可容许轨道后 $E(\mathcal{P})$ 的最小值. 那么 $\boldsymbol{e}\geqslant-\boldsymbol{m}$, 且 $\boldsymbol{e}=-\boldsymbol{m}$ 当且仅当对任意 $(i,j)\in\mathbb{S}\times\mathbb{S}_0$, 存在一条从 i 到 j 的可容许轨道 p, 使得 $\text{Elev(p)}=H(i)$. (参看后面的习题 5.6.5). 更一般地, 不论 $\boldsymbol{e}$ 取何值, 存在不依赖于 H 的常数 $0<c_-\leqslant c_+<\infty$, 使得对任意的 $\beta\geqslant 0$,*

$$c_-e^{-\beta(\boldsymbol{e}+\boldsymbol{m})}\leqslant\lambda_\beta\leqslant c_+e^{-\beta(\boldsymbol{e}+\boldsymbol{m})}.$$

证明: 因为当 H 用 $H-\boldsymbol{m}$ 来替代而 $\boldsymbol{e}$ 变作 $\boldsymbol{e}+\boldsymbol{m}$ 时, $\gamma(\beta)$ 和 $\boldsymbol{Q}(\beta)$ 都不改变, 因此我们可以假设 $\boldsymbol{m}=0$.

选取可容许轨道的一个集合 $\mathcal{P}=\{p(i,j):(i,j)\in\mathbb{S}^2\}$, 使得对于任意的 (i,j), $e(p(i,j))$ 等于 $e(p)$ 当 p 取遍所有从 i 到 j 的可容许轨

[4] 如果不是这样, 我们可以选取 k_0 为 H 达到最小值的点并且用 $H-H(k_0)$ 代替 H, 从而这样的条件满足. 这样的修改不改变 $\gamma(\beta)$, $\boldsymbol{Q}(\beta)$ 以及 (5.4.10) 右边的量的取值.

道后的最小值. 其次, 给定 $k_0 \in \mathbb{S}_0$, 由上面推理得

$$\boldsymbol{e} = \max_{i\in\mathbb{S}} e(p(i,k_0)).$$

特别地,由于 $e(p(i,k_0)) = \mathrm{Elev}(p(i,k_0)) - H(i) \geqslant 0,\ i \in \mathbb{S}$, 这就证明了 $\boldsymbol{e} \geqslant 0$, 并且显然有 $\boldsymbol{e} = 0$ 当且仅当对所有的 $i \in \mathbb{S}$, $H(i) = \mathrm{Elev}(p(i,k_0))$.

考虑 λ_β 的下界. 因为 $\boldsymbol{m} = 0$, 所以 $Z(\beta) \geqslant (\boldsymbol{\nu})_{k_0} > 0$, 再结合 (5.4.10) 式, 因此我们可取 $c_- = \dfrac{(\boldsymbol{\nu})_{k_0}}{W(\mathcal{P})}$.

最后, 为了证明上界, 选取 $\ell_0 \in \mathbb{S}\setminus\{k_0\}$, 使当 $p_0 \equiv p(\ell_0,k_0)$ 时, $e(p_0) = \boldsymbol{e}$, 令 Γ 为满足下述条件的 $i \in \mathbb{S}$ 的一个集合: 要么 $i = k_0$, 要么对从 i 到 k_0 的轨道 $p(i) \equiv p(i,k_0) \in \mathcal{P}$ 有 $\mathrm{Elev}(p(i)) < \mathrm{Elev}(p_0)$, 并且令 $f = \mathbf{1}_\Gamma$. 那么, 由于 $k_0 \in \Gamma$ 和 $\ell_0 \notin \Gamma$, 我们有

$$\begin{aligned}\mathrm{Var}_\beta(f) &= \left(\sum_{i\in\Gamma}(\gamma(\beta))_i\right)\left(\sum_{j\notin\Gamma}(\gamma(\beta))_j\right) \geqslant (\gamma(\beta))_{k_0})(\gamma(\beta))_{\ell_0})\\ &= \frac{(\boldsymbol{\nu})_{k_0}(\boldsymbol{\nu})_{\ell_0}}{Z(\beta)^2}e^{-\beta(H(k_0)+H(\ell_0))}.\end{aligned}$$

同时我们有

$$\begin{aligned}\mathcal{E}_\beta(f,f) &= \sum_{(i,j)\in\Gamma\times\Gamma^c}(\gamma(\beta))_i(\boldsymbol{Q}(\beta))_{ij}\\ &= \frac{1}{Z(\beta)}\sum_{(i,j)\in\Gamma\times\Gamma^c}(\boldsymbol{\nu})_i(\boldsymbol{A})_{ij}e^{-\beta(H(i)\vee H(j))}.\end{aligned}$$

但是若 $i \in \Gamma\setminus\{k_0\}, j \notin \Gamma$, $(\boldsymbol{A})_{ij} > 0$, 则 $H(j) \geqslant \mathrm{Elev}(p_0)$. 为了看出这一点, 考虑轨道 q, 它从 j 出发一步到达 i, 然后沿着 $p(i)$ 从 i 到达 k_0. 显然它是一条从 j 到 k_0 的可容许轨道. 因此 $\mathrm{Elev}(q) \geqslant \mathrm{Elev}(p(j)) \geqslant \mathrm{Elev}(p_0)$. 这意味着

$$\mathrm{Elev}(p_0) \leqslant \mathrm{Elev}(p(j)) \leqslant \mathrm{Elev}(q) = \mathrm{Elev}(p(i)) \vee H(j),$$

再结合 $\mathrm{Elev}(p(i)) < \mathrm{Elev}(p_0)$, 就得到结论 $H(j) \geqslant \mathrm{Elev}(p_0)$. 当 $j \notin \Gamma$ 且 $(\boldsymbol{A})_{k_0 j} > 0$ 时这一结论更容易证明, 因为在这种情形, (j,k_0) 是可容许轨道, 且

$$H(j) = \mathrm{Elev}((j,k_0)) \geqslant \mathrm{Elev}(p(j)) \geqslant \mathrm{Elev}(p_0).$$

因此, 将它代入前面 $\mathcal{E}_\beta(f,f)$ 的表达式, 得

$$\mathcal{E}_\beta(f,f) \leqslant \frac{e^{-\beta \mathrm{Elev}(p_0)}}{Z(\beta)} \sum_{(i,j)\in\Gamma\times\Gamma^c} (\boldsymbol{\nu})_i(\boldsymbol{A})_{ij},$$

由于 $e = \mathrm{Elev}(p_0) - H(k_0) - H(l_0)$, 上式意味着

$$\lambda_\beta \leqslant \frac{\mathcal{E}_\beta(f,f)}{\mathrm{Var}_\beta(f)} \leqslant \frac{Z(\beta)}{(\boldsymbol{\nu})_{k_0}(\boldsymbol{\nu})_{\ell_0}} \left(\sum_{(i,j)\in\Gamma\times\Gamma^c} (\boldsymbol{\nu})_i(\boldsymbol{A})_{ij} \right) e^{-\beta e}.$$

最后, 由于 $Z(\beta) \leqslant \|\boldsymbol{\nu}\|_v$, 上界得证. □

5.5 模拟退火

最后一节考察上一节中的思想的一个应用. 给定一个函数 $H : \mathbb{S} \to [0,\infty)$, 我们将叙述寻找 H 达到最小值点的一个方法, 这就是我们熟知的模拟退火或*Metropolies* 算法.

为了理解这一算法隐含的直观思想, 令 $\boldsymbol{A}$ 为 §5.4 所讨论的那种矩阵, 假设 0 是 H 的最小值, 记 $\mathbb{S}_0 = \{i : H(i) = 0\}$. 考虑这样的动态算法, 它引导你从任意一个初始点出发到达 $\mathbb{S}_0$, 途经相对于 $\boldsymbol{A}$ 是可容许的轨道 (即对轨道上相继的两点 k 和 ℓ, $(\boldsymbol{A})_{k\ell} > 0$). 其中一个是基于最速下降策略的. 即设某人在点 k, 如果至少有一点 ℓ 使得 $(\boldsymbol{A})_{k\ell} > 0$ 且 $H(\ell) \leqslant H(k)$, 他将移到满足这样条件的使得 $H(\ell)$ 达到最小的一点 ℓ; 如果对所有满足 $(\boldsymbol{A})_{k\ell} > 0$ 的 ℓ 都有 $H(\ell) > H(k)$, 他将原封不动. 正如 §5.4 末一样用隐喻来说, 只要避免陷到某 "山谷" 中, 这一算法非常有效. 这就是最速下降法是达到 H 的局部最小值的最有效策略的要点. 然而, 如果那个最小值不是全局最小, 那么, 通常你就上当受骗了! 因此, 如果要避免这样的不幸, 尽管你选择的是往山下爬, 有时候你不得不往上爬. 但是, 除非你对整个地形有详细的先验知识, 没有任何办法知道何时你要决定这样做. 由于这样的原因, 最好放弃唯理性的做法, 而随机地做出决定. 当然, 过一会儿以后, 你应该希望你将走出山谷, 也希望最速下降法变得愈加合理.

5.5.1. 算法: 为了尽可能地避免许多技术性的问题, 我们假设 $\mathbb{S}$ 有限且至少有两个元素. 下面令 $H : \mathbb{S} \to [0,\infty)$ 为我们要去确定其达到最小值点的函数, 并且不失一般性, 我们假设 0 为其最小值. 现取 $\boldsymbol{\nu}$ 使得 $(\boldsymbol{\nu})_i = 1$, $i \in \mathbb{S}$, 并取一个矩阵 $\boldsymbol{A}$, 使得 $(\boldsymbol{A})_{ii} = 0$ $(i \in \mathbb{S})$,

$(\boldsymbol{A})_{ij} = (\boldsymbol{A})_{ji} \geqslant 0\ (i \neq j)$ 且 $\boldsymbol{A}$ 在 $\mathbb{S}$ 上不可约 (可参看 (5.4.4)). 在实际应用中, $\boldsymbol{A}$ 的选取要使得当 $(\boldsymbol{A})_{ij} > 0$ 时 $H(j) - H(i)$ 的赋值 "尽可能地容易". 例如, 如果 $\mathbb{S}$ 有某种类型的自然邻域结构使得 $\mathbb{S}$ 是连通的, 并且当 j 与 i 相邻时 $H(j) - H(i)$ 的计算只需要很少的时间, 那么有理由取 $\boldsymbol{A}$ 使得 $(\boldsymbol{A})_{ij} = 0$ 除非 $j \neq i$ 与 i 相邻.

现在如 (5.4.2) 式那样定义 $\gamma(\beta)$, 如 (5.4.7) 式那样定义 $\boldsymbol{Q}(\beta)$. 显然, $\gamma(0)$ 为在 $\mathbb{S}$ 上标准化的均匀分布, 即 $(\gamma(0))_i = L^{-1}$, 其中 $L \equiv \#\mathbb{S} \geqslant 2$ 是 $\mathbb{S}$ 中元素的个数. 一方面, 当 β 变得越来越大时, $\gamma(\beta)$ 在 $\mathbb{S}_0$ 上变得越来越集中. 更确切地说, 因为 $Z(\beta) \geqslant \#\mathbb{S}_0 \geqslant 1$, 所以

$$\langle \mathbf{1}_{\mathbb{S}_0^c} \rangle_{\gamma(\beta)} \leqslant L e^{-\beta\delta}, \tag{5.5.1}$$

其中 $\delta \equiv \min\{H(j) : j \notin \mathbb{S}_0\}$. 另一方面, 当 β 变得越来越大时, 定理 5.4.11 告诉我们, 至少当 $\boldsymbol{e} > 0$ 时, $\lambda(\beta)$ 将会变得越来越小. 这样, 我们面临一对矛盾.

从引言所讨论的角度看, 这对取大 β 还是取小 β 的矛盾应该是预料之中的. 事实上, 取大 β 等价于采用一个近似的最速下降法, 而取小 β 等价于让点流动从而减少被陷进去的危险. 此外, 那个讨论的末尾给出了解决矛盾的建议. 也就是, 为了使得两者的优势达到最大, 我们应该从 $\beta = 0$ 出发, 允许 β 随时间增加[5]. 也即, 我们取 β 为一个递增的连续函数 $t \rightsquigarrow \beta(t)$, 满足 $\beta(0) = 0$. 为了不使我们的公式显得累赘, 我们引入如下记号: $Z(t) = Z(\beta(t))$, $\gamma_t = \gamma(\beta(t))$, $\langle\cdot\rangle_t = \langle\cdot\rangle_{\gamma_t}$, $\|\cdot\|_{2,t} = \|\cdot\|_{2,\gamma_t}$, $\mathrm{Var}_t = \mathrm{Var}_{\gamma_t}$, $\boldsymbol{Q}(t) = \boldsymbol{Q}(\beta(t))$, $\mathcal{E}_t = \mathcal{E}_{\beta(t)}$, $\lambda_t = \lambda_{\beta(t)}$.

因为在物理模型中, β 与温度的倒数成正比并且 β 随时间增加, 所以 $t \rightsquigarrow \beta(t)$ 称为*冷却方案*.

5.5.2. 转移概率的构造: 由于 Q-矩阵为时间相依的, 与之对应的*转移概率将是时间 - 非齐次的*. 因此, $t \rightsquigarrow Q(t)$ 确定的不是转移概率矩阵的一个单参数族, 而是对每个 $s \in [0, \infty)$, 它由*下述时间 - 非齐次 Kolmogorov 向前方程*确定一个从 $[s, \infty)$ 到转移概率矩阵的映射

[5] 实际上, 很有理由怀疑单调增加 β 是最好的途径. 事实上, "模拟退火" 的名称来自于下述思想, 就是人们要做的是去模拟化学家、材料科学家、熟练技工和 Metropolis 的追随者们所熟悉的退火过程. 也就是说, 这些人要做的是交替进行加热和冷却以达到他们的目标, 因此有理由相信我们应该跟随他们的思想. 然而, 基于不可宽恕但可理解的想法: 我的分析方法能够处理单调情形, 我选择不按照他们的思想.

$t \rightsquigarrow \boldsymbol{P}(s,t)$:

$$\frac{d}{dt}\boldsymbol{P}(s,t) = \boldsymbol{P}(s,t)\boldsymbol{Q}(t), \quad t \in (s,\infty), \quad \boldsymbol{P}(s,s) = \boldsymbol{I}. \tag{5.5.2}$$

尽管 (5.5.2) 式没有被包含在以前对 Kolmogorov 方程的分析中, 但是已经接近如此. 为了说明这一点, 我们通过一个逼近的方法求解方程 (5.5.2), 其中 (5.5.2) 式右边的 $\boldsymbol{Q}(t)$ 用下列矩阵代替:

$$\boldsymbol{Q}^{(N)}(t) \equiv \boldsymbol{Q}([t]_N),$$

其中 $[t]_N = \dfrac{m}{N}, t \in \left[\dfrac{m}{N}, \dfrac{m+1}{N}\right)$. 方程的解 $t \rightsquigarrow \boldsymbol{P}^{(N)}(s,t)$ 由 $\boldsymbol{P}^{(N)}(s,s) = \boldsymbol{I}$ 和

$$\boldsymbol{P}^{(N)}(s,t) = \boldsymbol{P}^{(N)}(s, s \vee [t]_N) e^{(t-s)\boldsymbol{Q}([t]_N)}, \quad t > s$$

给定. 由这一构造, 显然对每个 $N \geqslant 1$ 和 $t \geqslant s$, $\boldsymbol{P}^{(N)}(s,t)$ 为转移概率矩阵, 并且 $(s,t) \rightsquigarrow \boldsymbol{P}^{(N)}(s,t)$ 连续. 此外

$$\begin{aligned}
&\|\boldsymbol{P}^{(N)}(s,t) - \boldsymbol{P}^{(M)}(s,t)\|_{u,v} \\
\leqslant &\int_s^t \|\boldsymbol{Q}([\tau]_N) - \boldsymbol{Q}([\tau]_M)\|_{u,v} \|\boldsymbol{P}^{(N)}(s,\tau)\|_{u,v} d\tau + \\
&\int_s^t \|\boldsymbol{Q}([\tau]_M)\|_{u,v} \|\boldsymbol{P}^{(N)}(s,\tau) - \boldsymbol{P}^{(M)}(s,\tau)\|_{u,v} d\tau.
\end{aligned}$$

但是

$$\|\boldsymbol{Q}(\tau)\|_{u,v} \leqslant \|\boldsymbol{A}\|_{u,v},$$

$$\|\boldsymbol{Q}(\tau') - \boldsymbol{Q}(\tau)\|_{u,v} \leqslant \|\boldsymbol{A}\|_{u,v} \|H\|_u |\beta(\tau') - \beta(\tau)|,$$

从而

$$\begin{aligned}
&\|\boldsymbol{P}^{(N)}(s,t) - \boldsymbol{P}^{(M)}(s,t)\|_{u,v} \\
\leqslant &\|\boldsymbol{A}\|_{u,v} \|H\|_u \int_s^t |\beta([\tau]_N) - \beta([\tau]_M)| d\tau + \\
&\|\boldsymbol{A}\|_{u,v} \int_s^t |\boldsymbol{P}^{(N)}(\tau) - \boldsymbol{P}^{(M)}(\tau)| d\tau.
\end{aligned}$$

因此, 利用 Gronwall 不等式, 我们得

$$\begin{aligned}
&\sup_{0 \leqslant s \leqslant t \leqslant T} \|\boldsymbol{P}^{(N)}(s,t) - \boldsymbol{P}^{(M)}(s,t)\|_{u,v} \\
\leqslant &\|\boldsymbol{A}\|_{u,v} \|H\|_u e^{\|\boldsymbol{A}\|_{u,v} T} \int_0^T |\beta([\tau]_N) - \beta([\tau]_M)| d\tau.
\end{aligned}$$

由于 $\tau \rightsquigarrow \beta(\tau)$ 连续, 这就证明了 $\{\boldsymbol{P}^{(N)}(s,t) : N \geqslant 1\}$ 为 Cauchy 收敛序列, 即任取 $T > 0$,

$$\lim_{M\to\infty} \sup_{N\geqslant M} \sup_{0\leqslant s\leqslant t\leqslant T} \|\boldsymbol{P}^{(N)}(s,t) - \boldsymbol{P}^{(M)}(s,t)\|_{u,v} = 0.$$

因此, 我们知道存在一个连续映射 $(s,t) \rightsquigarrow \boldsymbol{P}(s,t)$, 使得关于 $\|\cdot\|_{u,v}$, $\boldsymbol{P}^{(N)}(s,t)$ 在有限区间上一致收敛到它. 特别地, 对每一 $t \geqslant s$, $\boldsymbol{P}(s,t)$ 为一个转移概率矩阵, 并且 $t \in [s,\infty) \mapsto \boldsymbol{P}(s,t)$ 为下述方程的一个连续解:

$$\boldsymbol{P}(s,t) = \boldsymbol{I} + \int_s^t \boldsymbol{P}(s,\tau)\boldsymbol{Q}(\tau)d\tau, \quad t \in [s,\infty),$$

这相当于 (5.5.2) 的积分形式. 进一步, 若 $t \in [s,\infty) \mapsto \boldsymbol{\mu}_t \in M_1(\mathbb{S})$ 连续可微, 则

$$\dot{\boldsymbol{\mu}}_t \equiv \frac{d}{dt}\boldsymbol{\mu}_t = \boldsymbol{\mu}_t\boldsymbol{Q}(t), \quad t \in [s,\infty) \Longleftrightarrow \boldsymbol{\mu}_t = \boldsymbol{\mu}_s\boldsymbol{P}(s,t), \quad t \in [s,\infty). \tag{5.5.3}$$

由于上述论断的充分性部分显然, 我们现在来证明必要性部分. 因此, 假设 $t \in [s,\infty) \mapsto \boldsymbol{\mu}_t \in \boldsymbol{M}_1(\mathbb{S})$ 满足 $\dot{\boldsymbol{\mu}}_t = \boldsymbol{\mu}_t\boldsymbol{Q}(t)$, 并记 $\boldsymbol{\omega}_t = \boldsymbol{\mu}_t - \boldsymbol{\mu}_s\boldsymbol{P}(s,t)$, 则

$$\boldsymbol{\omega}_t = \int_s^t \boldsymbol{\omega}(\tau)\boldsymbol{Q}_\tau d\tau,$$

从而

$$\|\boldsymbol{\omega}_t\|_{\boldsymbol{v}} \leqslant \|\boldsymbol{A}\|_{u,v} \int_s^t \|\boldsymbol{\omega}_\tau\|_{\boldsymbol{v}} d\tau.$$

因此, 再次利用 Gronwall 不等式, 知对一切 $t \geqslant s$ 有 $\boldsymbol{\omega}_t = \boldsymbol{0}$.

当然, 利用当 $\boldsymbol{\mu}_t = \boldsymbol{\delta}_i\boldsymbol{P}(s,t)$, $i \in \mathbb{S}$ 时的唯一性结果, 我们知 $(s,t) \rightsquigarrow \boldsymbol{P}(s,t)$ 是方程 (5.5.2) 的唯一解. 另外, 还得到下列时间 - 非齐次情形的 Chapman-Kolmogorov 方程:

$$\boldsymbol{P}(r,t) = \boldsymbol{P}(r,s)\boldsymbol{P}(s,t), \quad 0 \leqslant r \leqslant s \leqslant t. \tag{5.5.4}$$

事实上, 令 $\boldsymbol{\mu}_t = \boldsymbol{\delta}_i\boldsymbol{P}(r,t)$, $t \geqslant s$,注意到 $t \rightsquigarrow \boldsymbol{\mu}_t$ 满足 (5.5.3), 其中 $\boldsymbol{\mu}_s = \boldsymbol{\delta}_i\boldsymbol{P}(r,s)$. 我们得 $\boldsymbol{\mu}_t = \boldsymbol{\delta}_i\boldsymbol{P}(r,s)\boldsymbol{P}(s,t)$.

5.5.3. Markov 过程的描述: 给定一个概率向量 $\boldsymbol{\mu}$, 现在我们要构造一个 Markov 过程 $\{X(t) : t \geqslant 0\}$, 使得 $\boldsymbol{\mu}$ 为其初始分布, 而 $(s,t) \rightsquigarrow \boldsymbol{P}(s,t)$ 为其转移机制, 即

$$\mathrm{P}(X(0) = i) = (\boldsymbol{\mu})_i, \ \ \mathrm{P}(X(t) = j | X(\sigma), \sigma \in [0,s]) = \boldsymbol{P}(s,t)_{X(s)j}. \tag{5.5.5}$$

这里我们将要用到的思想基本上与在 §4.2.1 和 §2.1.1 中用到的相同. 然而, 由于时间 – 非齐次性, 致使我们有不可数个随机变量: 对每对 $(t,i)\in[0,\infty)\times\mathbb{S}$ 存在一对变量, 从而事情变得更加复杂. 为了处理这种情形, 不失一般性, 我们取 $\mathbb{S}=\{1,\cdots,L\}$, 并且对 $(t,i,j)\in[0,\infty)\times\mathbb{S}^2$, 记 $S(t,i,j)=\sum\limits_{\ell=1}^{j}e^{-\beta(t)(H(l)-H(i))^+}(\boldsymbol{A})_{i\ell}$, $\mathbb{S}(t,i,0)=0$. 定义

$$\Psi(t,i,u)=\begin{cases} j, & 若\ \dfrac{S(t,i,j-1)}{S(t,i,L)}\leqslant u<\dfrac{S(t,i,j)}{S(t,i,L)};\\ i, & 若\ u>1.\end{cases}$$

令 $\mathcal{T}:[0,\infty)\times\mathbb{S}\times[0,\infty)\to[0,\infty)$ 满足

$$\int_s^{s+\mathcal{T}(s,i,\xi)}S(\tau,i,L)d\tau=\xi.$$

下面, 设 X_0 为一个 $\mathbb{S}$ 值随机变量且其分布为 $\boldsymbol{\mu}$, 令 $\{E_n:n\geqslant 1\}$ 为一列独立单位指数随机变量且与 X_0 独立, 令 $\{U_n:n\geqslant 1\}$ 为一列 $[0,1)$ 上的独立均匀随机变量且与 $\sigma(\{X_0\}\bigcup\{E_n:n\geqslant 1\})$ 独立. 最后, 令 $J_0=0$, $X(0)=X_0$, 当 $n\geqslant 1$, 归纳地定义

$$J_n-J_{n-1}=\mathcal{T}(J_{n-1},X(J_{n-1}),E_n),\quad X(J_n)=\Psi(J_n,X(J_{n-1}),U_n),$$
$$X(t)=X(J_{n-1}),\quad J_{n-1}\leqslant t<J_n.$$

毋需大的变化, 把 §2.1.1 和 §4.2.2 中的论证结合起来, 可证 (5.5.5) 成立.

5.5.4. 冷却方案的选取: 在这一节中, 我们将给出一个有理数基, 在其上选取冷却方案 $t\rightsquigarrow\beta(t)$. 为此, 牢记我们要做什么是很要紧的. 我们试图让 Markov 过程 $\{X(t):t\geqslant 0\}$ 在下述意义下找出集合 $\mathbb{S}_0=\{j:H(j)=0\}$. 也就是说, 当 $t\to\infty$ 时, $\boldsymbol{P}(X(t)\notin\mathbb{S}_0)$ 趋于 0 的速度尽可能地快, 并且希望实现这一目标的方法使得 $X(t)$ 的分布尽可能地看起来像 γ_t. 因此, 一方面, 我们需要给 $\{X(t):t\geqslant 0\}$ 足够的时间以达到平衡, 使得 $X(t)$ 的分布看起来非常像 γ_t. 另一方面, 尽管可能会抑制平衡性, 但除非让 $\beta(t)$ 增长到无穷, 否则没有理由要使得 $X(t)$ 的分布看起来像 γ_t.

为了理解如何处理上面提到的内容, 令 $\boldsymbol{\mu}$ 为一个固定的初始分布, 令 $\boldsymbol{\mu}_t$ 为以 $\boldsymbol{\mu}$ 为初始分布的 Markov 过程 $\{X(t):t\geqslant 0\}$ 在 $t\geqslant 0$ 时的分布 (参看 §5.5.3). 等价地, $\boldsymbol{\mu}_t=\boldsymbol{\mu}\boldsymbol{P}(0,t)$, 其中 $\{\boldsymbol{P}(s,t):0\leqslant s\leqslant t<\infty\}$

为在 §5.5.2 中构造的转移概率矩阵族. 下面, 定义 $f_t : \mathbb{S} \to [0, \infty)$, 使得

$$f_t(i) = \frac{(\boldsymbol{\mu}_t)_i}{(\gamma_t)_i}, \quad t \geqslant 0, \quad i \in \mathbb{S}.$$

显然, f_t 的大小为衡量 $\boldsymbol{\mu}_t$ 与 γ_t 的相像程度提供了一个好的度量. 例如, 由 Schwarz 不等式和 (5.5.1) 式, 我们有

$$\begin{aligned} \mathrm{P}(X(t) \notin \mathbb{S}_0) &= \langle \mathbf{1}_{\mathbb{S}_0^c} \rangle_{\boldsymbol{\mu}_t} = \langle f_t \mathbf{1}_{\mathbb{S}_0^c} \rangle_t \\ &\leqslant \|f_t\|_{2,t} \sqrt{\langle \mathbf{1}_{\mathbb{S}_0^c} \rangle_t} \leqslant L^{\frac{1}{2}} \|f_t\|_{2,t} e^{-\frac{\beta(t)\delta}{2}}. \end{aligned} \tag{5.5.6}$$

因此如果能够让 $\|f_t\|_{2,t}$ 始终得到控制, 那么我们就已经前进了一步.

在有了上述讨论之后, 假设 $t \rightsquigarrow \beta(t)$ 是连续可微的. 注意到这保证了 $t \rightsquigarrow \|f_t\|_{2,t}^2$ 也是连续可微的. 事实上, 因为由 (5.4.3) 式得 $\dot{Z}(t) = -\dot{\beta}(t) Z(t) \langle H \rangle_t$, 所以

$$\begin{aligned} \frac{d}{dt} \|f_t\|_{2,t}^2 &= \frac{d}{dt} \left(\frac{1}{Z(t)} \sum_{i \in \mathbb{S}} (f_t(i))^2 e^{-\beta(t) H(i)} \right) \\ &= 2\langle f_t, \dot{f}_t \rangle_t - \dot{\beta}(t) \langle H - \langle H \rangle_t, f_t^2 \rangle_t. \end{aligned}$$

另一方面, 我们可以用另一种方式计算这一导数. 也就是, 任取函数 g, 由于 $\langle \boldsymbol{P}(0,t) g \rangle_{\boldsymbol{\mu}} = \langle g \rangle_{\boldsymbol{\mu}_t} = \langle f_t, g \rangle_t$, 所以

$$\|f_t\|_{2,t}^2 = \langle f \rangle_{\boldsymbol{\mu}_t} = \langle \boldsymbol{P}(0,t) f_t \rangle_{\boldsymbol{\mu}},$$

再由 (5.5.2) 式, 得

$$\frac{d}{dt} \|f_t\|_{2,t}^2 = \langle \boldsymbol{P}(0,t) \boldsymbol{Q}(t) f_t \rangle_{\boldsymbol{\mu}} + \langle \boldsymbol{P}(0,t) \dot{f}_t \rangle_{\boldsymbol{\mu}} = -\mathcal{E}_t(f_t, f_t) + \langle f_t, \dot{f}_t \rangle_t.$$

因此, 综合上述各式并消去含有 $\dot{f}_t$ 的项, 得

$$\begin{aligned} \frac{d}{dt} \|f_t\|_{2,t}^2 &= -2\mathcal{E}_t(f_t, f_t) + \dot{\beta}(t) \langle H - \langle H \rangle_t, f_t^2 \rangle_t \\ &\leqslant -(2\lambda_t - \|H\|_u \dot{\beta}(t)) \|f_t\|_{2,t}^2 + 2\lambda_t, \end{aligned}$$

其中, 在第二行我们利用了 $\langle f_t \rangle_t = 1$ 因此 $\mathrm{Var}_t(f) = \|f_t\|_{2,t}^2 - 1$ 的事实. 于是综上所述得证

$$\|H\|_u \dot{\beta}(t) \leqslant \lambda_t \Longrightarrow \frac{d}{dt} \|f_t\|_{2,t}^2 \leqslant -\lambda_t \|f_t\|_{2,t}^2 + 2\lambda_t.$$

容易对上述关于 $\|f_t\|_{2,t}^2$ 的微分不等式求积分. 即

$$\frac{d}{dt}(e^{\Lambda(t)}\|f_t\|_{2,t}^2) \leqslant 2\lambda_t e^{\Lambda(t)}, \quad \text{其中 } \Lambda(t) = \int_0^t \lambda_\tau d\tau.$$

因此

$$\|f_t\|_{2,t}^2 \leqslant e^{-\Lambda(t)}\|f_0\|_{2,0}^2 + 2(1 - e^{-\Lambda(t)}) \leqslant \|f_0\|_{2,0}^2 \vee 2.$$

此外, 由于 $(\gamma_0)_i = L^{-1}$, 其中 $L = \#\mathbb{S} \geqslant 2$, 所以 $\|f_0\|_{2,0}^2 \leqslant L$, 从而

$$\|H\|_u \dot{\beta}(t) \leqslant \lambda_t \Longrightarrow \|f_t\|_{2,t} \leqslant L^{\frac{1}{2}}. \tag{5.5.7}$$

最后将找出如何选取 $t \rightsquigarrow \beta(t)$ 的方法, 使得它满足 (5.5.7) 中的条件; 当然, 我们只对 $\mathbb{S}_0 \neq \mathbb{S}$ 的情形, 即等价地, $\|H\|_u > 0$ 的情形, 感兴趣. 由定理 5.4.11, 我们知道 $\lambda_t \geqslant c_- e^{-\beta(t)\boldsymbol{e}}$. 因此, 我们可取

$$\beta(t) = \begin{cases} \dfrac{1}{\boldsymbol{e}} \log\left(1 + \dfrac{c_- \boldsymbol{e} t}{\|H\|_u}\right), & \text{当 } \boldsymbol{e} > 0; \\ \dfrac{c_- t}{\|H\|_u}, & \text{当 } \boldsymbol{e} = 0. \end{cases} \tag{5.5.8}$$

其中 $\boldsymbol{e} = 0$ 情形的取值是通过明显的取极限过程得到的. 把这些和 (5.5.7), (5.5.6) 结合起来, 我们证明了当 $\beta(t)$ 由 (5.5.8) 式给定时,

$$\mathrm{P}(X(t) \notin \mathbb{S}_0) \leqslant L \begin{cases} \left(1 + \dfrac{c_- \boldsymbol{e} t}{\|H\|_u}\right)^{-\frac{\delta}{2\boldsymbol{e}}}, & \text{当 } \boldsymbol{e} > 0; \\ e^{-\frac{\delta c_- t}{2\|H\|_u}}, & \text{当 } \boldsymbol{e} = 0. \end{cases} \tag{5.5.9}$$

注: 当 $\boldsymbol{e} = 0$ 时, (5.5.9) 式的结论值得进一步讨论. 特别地, 应该注意到, $\boldsymbol{e} = 0$ 并不能保证最速下降策略的成功. 事实上, $\boldsymbol{e} = 0$ 仅仅意味着每个 i 可以通过一条可容许轨道与 $\mathbb{S}_0$ 相连, 沿着这条轨道 H 单调非增 (参看习题 5.6.5), 它并不排除这样的可能性, 即当采用最速下降法时, 会选取差的轨道从而上当受骗. 因此, 即使在这种情形, 我们仍需要足够的随机性来到处搜寻直到找到好的轨道.

5.5.5. 小的改进: 一个只在 $\boldsymbol{e} > 0$ 时才真正有意义的事实是, 我们可以使用同样的分析来控制 $\|f_t\|_{q,t} \equiv (\langle |f_t|^q \rangle_t)^{\frac{1}{q}}$, $q \in [2, \infty)$. 结果是, 我们可以证明, 对每一个 $\theta \in (0,1)$, 存在一个冷却方案使得 $\boldsymbol{P}(X(t) \notin \mathbb{S}_0)$ 至少以 $t^{-\frac{\theta\delta t}{\boldsymbol{e}}}$ 这么快的速度趋于 0. θ 和 q 的关系由 $\theta = 1 - \dfrac{1}{q}$ 给出.

为了证明这一点, 我们首先计算 $\frac{d}{dt}\|f_t\|_{q,t}^q$ 两次, 各利用下述表达式一次:

$$\|f_t\|_{q,t}^q = \langle f_t^q\rangle_t, \quad \|f_t\|_{q,t}^q = \langle \boldsymbol{P}(0,t)f_t^{q-1}\rangle_{\boldsymbol{\mu}}.$$

然后, 消去所得方程式中的 $\dot{f}_t$, 从而得到

$$\frac{d}{dt}\|f_t\|_{q,t}^q = -q\mathcal{E}_t(f_t, f_t^{q-1}) - \frac{\dot{\beta}(t)}{q'-1}\langle f_t^q, H - \langle H\rangle_t\rangle_t,$$

其中 $q' = \frac{q}{q-1}$,

$$\begin{aligned}\mathcal{E}_t(\varphi,\psi) &\equiv \frac{1}{2}\sum_{i\neq j}(\gamma_t)_i(\boldsymbol{Q}(t))_{ij}(\varphi(j)-\varphi(i))(\psi(j)-\psi(i))\\ &= -\langle\varphi, \boldsymbol{Q}(t)\psi\rangle_t.\end{aligned}$$

至此, 我们只要证明

$$\mathcal{E}_t(f_t, f_t^{q-1}) \geqslant \frac{4(q-1)}{q^2}\mathcal{E}_t(f_t^{\frac{q}{2}}, f_t^{\frac{q}{2}}),$$

而稍作思考就会明白为证这个不等式只要验证对任意一对 $(a,b) \in [0,\infty)^2$, 有

$$(b^{\frac{q}{2}} - a^{\frac{q}{2}})^2 \leqslant \frac{q^2}{4(q-1)}(b-a)(b^{q-1}-a^{q-1}),$$

这一不等式可由微积分基本定理和 Schwarz 不等式得到. 因此, 结合上述和 (5.3.6) 式, 得

$$\frac{d}{dt}\|f_t\|_{q,t}^q \leqslant -\frac{1}{q'}(4\lambda_t - q\dot{\beta}(t)\|H\|_u)\|f_t\|_{q,t}^q + \frac{4\lambda_t}{q'}\langle f_t^{\frac{q}{2}}\rangle_t^2.$$

为了继续进一步的讨论, 我们必须掌握如何根据 $\|f_t\|_{q,t}^q$ 来控制 $\langle f_t^{\frac{q}{2}}\rangle_t^2$. 当 $q=2$ 时, 无需证明, 因为这时两者相等从而控制系数为 1. 当 $q>2$ 时, 我们不再有这样的控制. 不过, 由 $\langle f_t^{\frac{q}{2}}\rangle_t^2 = \langle f_t^{\frac{q}{2}-1}\rangle_{\boldsymbol{\mu}_t}^2$, 再利用习题 5.6.2 (c), 我们有

$$\langle f_t^{\frac{q}{2}-1}\rangle_{\boldsymbol{\mu}_t}^2 \leqslant \langle f_t^{q-1}\rangle_{\boldsymbol{\mu}_t}^{\frac{q-2}{q-1}} = \langle f_t^q\rangle_t^{\frac{q-2}{q-1}},$$

得到 $\langle f_t^{\frac{q}{2}}\rangle_t^2 \leqslant \langle f_t^q\rangle_t^{\frac{q-2}{q-1}}$. 借助于这一估计的帮助, 我们得到如下的微分不等式:

$$\frac{d}{dt}\|f_t\|_{q,t}^q \leqslant -\frac{1}{q'}\left(4\lambda_t - q\|H\|_u\dot{\beta}(t)\right)\|f_t\|_{q,t}^q + \frac{4\lambda_t}{q'}\left(\|f_t\|_{q,t}^q\right)^{1-\frac{q'}{q}}.$$

最后, 取 $\beta(t)=\frac{1}{e}\log\left(1+\frac{3c_-et}{q\|H\|_u}\right)$, 上述不等式可由下式代替:

$$\frac{d}{dt}\|f_t\|_{q,t}^q\leqslant-\frac{\lambda_t}{q'}\|f_t\|_{q,t}^q+\frac{4\lambda_t}{q'}\left(\|f_t\|_{q,t}^q\right)^{1-\frac{q'}{q}}.$$

通过积分, 由上述不等式可得到

$$\|f_t\|_{q,t}\leqslant 2^{\frac{1}{q'}}\vee\|f_0\|_{q,0}.$$

注: 上述推导过程基本上是 Poincaré 不等式的推论. 实际上, 如果把这些推导和一般性的 Sobolev 不等式的分析思想结合起来, 有可能得到更好的估计. 有兴趣的读者可以参看 [3]. 本节的内容来源于该参考文献.

5.6 习题

5.6.1. 称函数 $\psi:[0,\infty)\to\mathbb{R}$ 是凸的, 如果 ψ 的任意两点间的曲线在连接这两点的割线的下方, 即满足

$$\psi\big((1-\theta)s+\theta t\big)\leqslant(1-\theta)\psi(s)+\theta\psi(t),\quad 0\leqslant s<t,\quad \theta\in[0,1].\quad (*)$$

这个习题考察凸函数的各种性质. 这些性质都由凸函数斜率的单调非降性导出.

(**a**) 设 $\{\psi_n\}_1^\infty\bigcup\{\psi\}$ 是 $[0,\infty)$ 上的函数, 且对每个 $t\in[0,\infty)$ 有 $\psi_n(t)\to\psi(t)$. 证明: 如果每个 ψ_n 是非增的, 则 ψ 也是非增的; 如果每个 ψ_n 是凸的, 则 ψ 也是凸的.

(**b**) 设 $\psi:[0,\infty)\to\mathbb{R}$ 连续且在 $(0,\infty)$ 上两次连续可微. 证明 ψ 在 $[0,\infty)$ 上是凸的当且仅当在 $(0,\infty)$ 上 $\ddot{\psi}\geqslant 0$.

提示: "必要性" 由下式推出

$$\ddot{\psi}(t)=\lim_{h\searrow 0}\frac{\psi(t+h)+\psi(t-h)-2\psi(t)}{h^2},\quad t\in(0,\infty).$$

为证明 "充分性", 给定 $0<s<t$, 对 $\epsilon>0$ 记

$$\varphi_\epsilon(\theta)=\psi\big((1-\theta)s+\theta t\big)-(1-\theta)\psi(s)-\theta\psi(t)-\epsilon\theta(1-\theta),\quad\theta\in[0,1].$$

注意到 $\varphi_\epsilon(0)=0=\varphi_\epsilon(1)$, 且在 $(0,1)$ 上 $\ddot{\varphi}_\epsilon>0$. 从而由二阶导数判别法, φ_ϵ 在 $(0,1)$ 上不能达到最大值. 最后令 $\epsilon\searrow 0$.

(c) 设 ψ 在 $[0,\infty)$ 是凸的. 证明: 对每个 $s \in [0,\infty)$,

$$t \in (s,\infty) \mapsto \frac{\psi(s)-\psi(t)}{t-s}$$

是非增的.

(d) 设 ψ 在 $[0,\infty)$ 是凸的, $0 \leqslant s < t \leqslant u < w$. 证明

$$\frac{\psi(s)-\psi(t)}{t-s} \geqslant \frac{\psi(u)-\psi(w)}{w-u}.$$

提示: 归纳到 $u=t$ 的情形.

5.6.2. 给定一个概率向量 $\boldsymbol{\mu} \in [0,1]^{\mathbb{S}}$, 有许多方法证明对任何 $f \in L^2(\boldsymbol{\mu})$ 有 $\langle f\rangle_{\boldsymbol{\mu}}^2 \leqslant \langle f^2\rangle_{\boldsymbol{\mu}}$. 例如, 可以利用 Schwarz 不等式 $|\langle f,g\rangle_{\boldsymbol{\mu}}| \leqslant \|f\|_{2,\boldsymbol{\mu}}\|g\|_{2,\boldsymbol{\mu}}$ 并令 $g=1$ 得到这个不等式. 另一个方法是, 利用 $0 \leqslant \mathrm{Var}_{\boldsymbol{\mu}}(f) = \langle f^2\rangle_{\boldsymbol{\mu}} - \langle f\rangle_{\boldsymbol{\mu}}^2$. 然而这些证明都没有揭示凸性所起的本质作用. 也就是说, 这个题的目的是证明: 对任意的非降、连续、凸函数 $\psi:[0,\infty)\to[0,\infty)$ 和任意的 $f:\mathbb{S}\to[0,\infty)$, 有

$$\psi\big(\langle f\rangle_{\boldsymbol{\mu}}\big) \leqslant \langle \psi\circ f\rangle_{\boldsymbol{\mu}}, \tag{5.6.3}$$

其中当 $\langle f\rangle_{\boldsymbol{\mu}} = \infty$ 时, 上式由 $\psi(\infty) \equiv \lim\limits_{t\nearrow\infty}\psi(t)$ 给定. 不等式 (5.6.3) 是更一般的 *Jensen 不等式* (参看 [8] 的定理 6.1.1) 的特例.

(a) 对 $n \geqslant 2$ 进行归纳证明

$$\psi\left(\sum_{k=1}^n \theta_k x_k\right) \leqslant \sum_{k=1}^n \theta_k \psi(x_k)$$

对所有的 $(\theta_1,\cdots,\theta_n)\in[0,1]^n$, $\sum\limits_{k=1}^n \theta_k = 1$ 和 $(x_1,\cdots,x_n)\in[0,\infty)^n$ 成立.

(b) 设一列有限集 $\{F_N\}_1^\infty$ 为 $\mathbb{S}$ 的单调不减穷举 (即 $\bigcup\limits_{N=1}^\infty F_N = \mathbb{S}$), 满足 $\boldsymbol{\mu}(F_N) \equiv \sum\limits_{i\in F_N}(\boldsymbol{\mu})_i > 0$. 利用 (a) 证明, 对每个 N,

$$\psi\left(\sum_{i\in F_N} f(i)(\boldsymbol{\mu})_i\right) \leqslant \frac{\sum\limits_{i\in F_N}\psi\big(f(i)\big)(\boldsymbol{\mu})_i}{\mu(F_N)} \leqslant \frac{\langle\psi\circ f\rangle_{\boldsymbol{\mu}}}{\mu(F_N)},$$

然后令 $N\to\infty$ 得到要证明的结论.

(c) 利用 (5.6.3) 式证明, 对任意的 $0<p\leqslant q<\infty$ 和 $f:\mathbb{S}\to[0,\infty)$ 有 $\langle f^p\rangle_{\boldsymbol{\mu}}^{\frac{1}{p}} \leqslant \langle f^q\rangle_{\boldsymbol{\mu}}^{\frac{1}{q}}$.

5.6.4. Gronwall 不等式有很多形式, 最基本的表述为, 如果 $u:[0,T]\to[0,\infty)$ 是连续函数, 满足

$$u(t)\leqslant A+B\int_0^t u(\tau)d\tau,\ \ t\in[0,T],$$

则 $u(t)\leqslant Ae^{Bt}, t\in[0,T]$. 证明这种形式的 *Gronwall* 不等式.

提示: 记 $U(t)=\int_0^t u(\tau)d\tau$, 证明 $\dot{U}(t)\leqslant A+BU(t)$. 由此得到 $U(t)\leqslant\dfrac{A}{B}(e^{Bt}-1)$.

5.6.5. 设 $\boldsymbol{e},\boldsymbol{m}$ 如定理 5.4.11 那样定义, 证明 $\boldsymbol{e}=-\boldsymbol{m}$ 的充分必要条件是对任意的 $(i,j)\in\mathbb{S}\times\mathbb{S}_0$, 存在一条从 i 出发在 j 处结束的可容许轨道 $(i_0,\cdots,i_n)$, 使得 H 沿着它单调非增. 也即, $i_0=i$, $i_n=j$, 并且对每个 $1\leqslant m\leqslant n$, $\boldsymbol{A}_{i_{m-1},i_m}>0$ 和 $H(i_m)\leqslant H(i_{m-1})$.

5.6.6. 该题涉及 §5.1.3 的内容, 说明除了 (5.1.10) 式以外, 那一节的所有内容或多或少可以推广到一般的不可约、正常返的转移概率 $\boldsymbol{P}$, 而不管它是不是可逆的. 和以前一样, 令 $\boldsymbol{\pi}=\boldsymbol{\pi}^{\mathbb{S}}$ 为唯一的 $\boldsymbol{P}$-平稳概率向量. 另外, 为了下面的习题, 考察满足 $|f|\in L^2(\boldsymbol{\pi})$ 的函数 $f:\mathbb{S}\to\mathbb{C}$ 组成的空间 $L^2(\boldsymbol{\pi},\mathbb{C})$.

(a) 对 (5.1.2) 定义的 $\boldsymbol{P}^\top$, 证明 $1\geqslant(\boldsymbol{P}\boldsymbol{P}^\top)_{ii}=(\boldsymbol{\pi})_i\sum\limits_{j\in\mathbb{S}}\dfrac{(\boldsymbol{P})_{ij}^2}{(\boldsymbol{\pi})_j}$, 并证明定义 $\boldsymbol{P}f(i)\equiv\sum\limits_{j\in\mathbb{S}}f(j)(\boldsymbol{P})_{ij}$ 中的级数对任意的 $i\in\mathbb{S}$ 和 $f\in L^2(\boldsymbol{\pi},\mathbb{C})$ 绝对收敛.

(b) 证明 $\|\boldsymbol{P}f\|_{2,\boldsymbol{\pi}}\leqslant\|f\|_{2,\boldsymbol{\pi}}$, $f\in L^2(\boldsymbol{\pi},\mathbb{C})$, 并把 (5.1.8) 和 (5.1.9) 推广到现在的情形, 也即取其中的 $f\in L^2(\boldsymbol{\pi},\mathbb{C})$.

5.6.7. 继续习题 5.6.6 的讨论, 我们现在说明可逆性在 §5.1.4 中只起微小的作用. 令 $\boldsymbol{P}$ 为 $\mathbb{S}$ 上不可约的转移概率且是正常返的, 令 d 是它的周期, 记 $\theta_d=e^{\sqrt{-1}2\pi d^{-1}}$.

(a) 证明对每个 $0\leqslant m<d$ 存在一个函数 $f_m:\mathbb{S}\to\mathbb{C}$ 使得 $|f_m|\equiv1$ 且 $\boldsymbol{P}f_m=\theta_d^m f_m$.

提示: 如同 §3.2.7, 取一个循环分解 $(\mathbb{S}_0,\cdots,\mathbb{S}_{d-1})$, 并利用 (3.2.19) 式.

(b) 给定 $\alpha\in\mathbb{R}\backslash\{0\}$, 记 $\theta_\alpha=e^{\sqrt{-1}2\pi\alpha^{-1}}$. 证明存在满足 $\boldsymbol{P}f=\theta_\alpha f$ 的 $f\in L^2(\boldsymbol{\pi},\mathbb{C})\backslash\{\boldsymbol{0}\}$ 当且仅当对某个 $m\in\boldsymbol{Z}\backslash\{0\}$, $d=m\alpha$.

提示: 由 (a), 只要证明, 除非 $d=m\alpha$, 其中 m 为非零整数, 否则这样的 f 不存在. 因此, 设这样的 f 对某个 α 存在, 而 α 不是 $\dfrac{d}{m}$ 形式的

有理数, 取 $i \in \mathbb{S}$ 使得 $f(i) \neq 0$, 得到与 $\lim_{n\to\infty} \boldsymbol{P}^{nd} f(i)$ 存在这一事实相矛盾的结论.

(c) 设 $f \in L^2(\mathbb{S}, \mathbb{C})$ 是对某个 $m \in \mathbb{Z}$, $\boldsymbol{P}f = \theta_d^m f$ 的非平凡、有界解, 令 $(\mathbb{S}_0, \cdots, \mathbb{S}_{d-1})$ 为 $\mathbb{S}$ 的一个循环分解. 证明对每个 $0 \leqslant r < d$, $f \restriction \mathbb{S}_r \equiv \theta_d^{-rm} c_0$, 其中 $c_0 \in \mathbb{C}\backslash\{0\}$. 特别地, 最多差一个常数因子, 对每个 $0 \leqslant m < d$ 确实存在一个非平凡、有界的 $f \in L^2(\mathbb{S}, \mathbb{C})$ 使得 $\boldsymbol{P}f = \theta_d^m f$.

(d) 令 H 是由满足 $\boldsymbol{P}f = \theta f$ (对某个 $\theta \in \mathbb{C}$, $|\theta| = 1$) 的 $f \in L^2(\mathbb{S}, \mathbb{C})$ 组成的线性子空间. 结合 (b) 和 (c), 证明 d 是 $\mathbb{C}$ 上的向量空间 H 的维数.

(e) 假设对每对 $(i,j) \in \mathbb{S}^2$, $(\boldsymbol{P})_{ij} > 0 \implies (\boldsymbol{P})_{ji} > 0$. 证明 $d \leqslant 2$, 并且 $d = 2$ 的充分必要条件是存在非平凡、有界的 $f : \mathbb{S} \to \mathbb{R}$, 使得 $\boldsymbol{P}f = -f$. 因此, 这是可逆转移概率矩阵的、在定理 5.1.14 中起本质作用的唯一性质.

5.6.8. 本题为考察非负定性和非周期性之间的关系提供了另一种思考. 令 $\boldsymbol{P}$ 为 $\mathbb{S}$ 上一个转移概率矩阵, 不一定是不可约的. 设 $\boldsymbol{\mu}$ 是一个概率向量, 满足逐一平衡条件 $(\boldsymbol{\mu})_i(\boldsymbol{P})_{ij} = (\boldsymbol{\mu})_j(\boldsymbol{P})_{ji}$, $(i,j) \in \mathbb{S}^2$. 进一步, 假设 $\boldsymbol{P}$ 在 $L^2(\boldsymbol{\mu})$ 上是非负定的, 即对任意的有界函数 $f : \mathbb{S} \to \mathbb{R}$, 有 $\langle f, \boldsymbol{P}f\rangle_{\boldsymbol{\mu}} \geqslant 0$. 证明 $(\boldsymbol{\mu})_i > 0 \implies (\boldsymbol{P})_{ii} > 0$, 从而如果 $(\boldsymbol{\mu})_i > 0$, 则 i 是非周期的. 按以下步骤可以得到所要证的结论.

(a) 定义矩阵 $\boldsymbol{A}$ 使得 $(\boldsymbol{A})_{ij} = \langle \mathbf{1}_{\{i\}}, \boldsymbol{P}\mathbf{1}_{\{j\}}\rangle_{\boldsymbol{\mu}}$, 证明 $\boldsymbol{A}$ 是对称的 (即 $(\boldsymbol{A})_{ij} = (\boldsymbol{A})_{ji}$), 并且是非负定的, 即对任意只有有限个非零元素的 $\boldsymbol{x} \in \mathbb{R}^{\mathbb{S}}$ 有 $\sum_{ij} (\boldsymbol{A})_{ij}(\boldsymbol{x})_i(\boldsymbol{x})_j \geqslant 0$.

(b) 给定 $i \neq j$, 考察 $L^2(\boldsymbol{\pi})$ 中的平面 $\{\alpha \mathbf{1}_i + \beta \mathbf{1}_j : \alpha, \beta \in \mathbb{R}\}$, 利用推导 (5.1.5) 的方法证明 $(\boldsymbol{A})_{ij}^2 \leqslant (\boldsymbol{A})_{ii}(\boldsymbol{A})_{jj}$. 特别地, 若 $(\boldsymbol{A})_{ii} = 0$, 则对任意的 $j \in \mathbb{S}$, $(\boldsymbol{A})_{ij} = 0$.

(c) 注意到 $\sum_{j\in\mathbb{S}} (\boldsymbol{A})_{ij} = (\boldsymbol{\mu})_i$, 完成证明.

(d) 之后, 再证明并不需要假定 $\boldsymbol{P}$ 是非负定的, 只需要假定对给定的 $i \in \mathbb{S}$, 每个子矩阵

$$
\begin{pmatrix}
\langle \mathbf{1}_{\{i\}}, \boldsymbol{P}\mathbf{1}_{\{i\}}\rangle_{\boldsymbol{\mu}} & \langle \mathbf{1}_{\{i\}}, \boldsymbol{P}\mathbf{1}_{\{j\}}\rangle_{\boldsymbol{\mu}} \\
\langle \mathbf{1}_{\{j\}}, \boldsymbol{P}\mathbf{1}_{\{i\}}\rangle_{\boldsymbol{\mu}} & \langle \mathbf{1}_{\{j\}}, \boldsymbol{P}\mathbf{1}_{\{j\}}\rangle_{\boldsymbol{\mu}}
\end{pmatrix}
$$

是非负定的.

5.6.9. 设 $\boldsymbol{P}$ 是不可约的、正常返的转移概率矩阵, 其平稳分布为 $\boldsymbol{\pi}$. 回顾 (5.1.2) 式, 证明 (5.1.3) 式中的第一个是不可约的, 而第二个不一定是不可约的. 再证明第一个的周期不会超过 $\boldsymbol{P}$ 的周期, 而第二个总是非负定的. 因此, 如果 $\boldsymbol{P}^{\top}\boldsymbol{P}$ 是不可约的, 则它必是非周期的.

5.6.10. 在许多应用中, Markov 链的构造由一个有限的简单图 $(\mathbb{S}, E)$ 决定. 更确切地说, 状态空间 $\mathbb{S}$ 是图的顶点集合, $\mathbb{S}^2$ 的对称子集 E 是图的边集, 简单性是指没有回路和双重边. 下面, 我们称 j 是 i 的最近邻点, 记做 $j \in \mathcal{N}(i)$, 当且仅当 $(i,j) \in E$. 我们始终假设对每个 $i \in \mathbb{S}$, $d(i) = \#\mathcal{N}(i)$ 是正的. 与 $(\mathbb{S}, E)$ 相应的转移概率矩阵 $\boldsymbol{P}$ 定义为 $(\boldsymbol{P})_{ij} = \dfrac{1}{d(i)}$, $j \in \mathcal{N}(i)$, 否则为零. 有兴趣的读者可在 [1] 中看到更多的例子.

(a) 证明 $\boldsymbol{P}$ 是不可约的当且仅当图是连通的, 即每对 $(i,j) \in \mathbb{S}^2$ 存在相应的一个有限序列 $(i_0, \cdots, i_n) \in \mathbb{S}^{n+1}$, 使得 $i = i_0$, $j = i_n$, 且对 $1 \leqslant m \leqslant n$, $(i_{m-1}, i_m) \in E$. 进一步, 在不可约性假设之下, 证明 Markov 链是非周期的当且仅当图不是二分的, 即不存在非空的子集 $\mathbb{S}' \subsetneqq \mathbb{S}$, 使得每一条边将 $\mathbb{S}'$ 中的一点连接到 $\mathbb{S} \backslash \mathbb{S}'$ 中的一点.

(b) 由 $(\boldsymbol{\pi})_i = \dfrac{d(i)}{2\#E}$ 定义概率向量 $\boldsymbol{\pi}$, 证明 $\boldsymbol{\pi}$ 关于 $\boldsymbol{P}$ 是可逆的.

(c) 假设图是连通的, 如 §5.2.2 中取一个可容许轨道集 $\mathcal{P} = \{p(i,j) : (i,j) \in \mathbb{S}^2 \backslash D\}$, 证明

$$\beta_+ \geqslant \frac{2\#E}{D^2 L(\mathcal{P}) B(\mathcal{P})}, \tag{5.6.11}$$

其中 $D \equiv \max\limits_{i \in \mathbb{S}} d(i)$, $L(\mathcal{P})$ 是 $\mathcal{P}$ 中的轨道最大长度 (即边数), $B(\mathcal{P})$ 是*瓶颈系数*, 即经过一条边 $e \in E$ 之轨道 $p \in \mathcal{P}$ 的最大数目. 显然, 如果选取由测地线 (连接所有端点的长度最短的轨道) 组成的 $\mathcal{P}$, 那么 $L(\mathcal{P})$ 就是 $(\mathbb{S}, E)$ 的直径, 并且达到了尽可能的小. 另一方面, 因为它可能产生坏的瓶颈 (许多轨道通过一给定的边), 选取的测地线可能不是最优的.

(d) 假设图是连通的并且不是二分的, 取 $\mathcal{P} = \{p(i) : i \in \mathbb{S}\}$ 为一个长度为奇数的可容许的闭轨道的集合, 证明

$$\beta_- \geqslant \frac{2}{D L(\mathcal{P}) B(\mathcal{P})}. \tag{5.6.12}$$

5.6.13. 这是一个利用习题 5.6.10 给出接近最优结果的例子. 令 $N \geqslant 2$, 考察由复平面 $\mathbb{C}$ 上 N 个单位根组成的集合 $\mathbb{S}$. 令 E 为相邻单位根对

所组成的集合, 即 $(e^{\frac{\sqrt{-1}2\pi(m-1)}{N}}, e^{\frac{\sqrt{-1}2\pi m}{N}})$ 组成的集合. 最后, 如上面定义 $\boldsymbol{P}$ 和 $\boldsymbol{\pi}$.

(a) 作为上面的 (c) 的应用, 证明 $\beta_+ \leqslant 1 - \dfrac{8N}{(N-1)^2(N+1)}$.

(b) 假设 N 为奇数, 利用上面的 (d), 证明 $\beta_- \geqslant -1 + \dfrac{1}{N^2}$.

5.6.14. 在本题中, 我们给出一类可逆 Markov 过程的大概介绍, 这类 Markov 过程为某类物理系统提供了有些理想化的数学模型. 在文献中, 这类物理系统被称为*自旋 - 反转* (*spin-flip*) *系统*, 这样叫的理由很快就会清楚, 他们是 Glauber 动力系统最早期的例子之一. 这里状态空间 $\mathbb{S} = \{-1,1\}^N$ 被认为是一个有 N 个粒子的系统的配置空间, 每一个配置有自旋状态 ("spin") +1 或 −1. 为了更方便, 我们用 $\boldsymbol{\omega} = (\omega_1, \cdots, \omega_N)$ 或 $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_N)$ 表示 $\mathbb{S}$ 中的一般元素. 给定 $\boldsymbol{\omega}$ 和 $1 \leqslant k \leqslant N$, $\hat{\boldsymbol{\omega}}^k$ 表示将 $\boldsymbol{\omega}$ 的第 k 个自旋状态反转得到的配置, 即 $\hat{\boldsymbol{\omega}}^k = (\omega_1, \cdots, \omega_{k-1}, -\omega_k, \omega_{k+1}, \cdots, \omega_N)$. 下面对定义在 $\mathbb{S}$ 上的 $\mathbb{R}$ 值函数 f 和 g, 定义

$$\Gamma(f,g)(\boldsymbol{\omega}) = \sum_{k=1}^{N} \big(f(\hat{\boldsymbol{\omega}}^k) - f(\boldsymbol{\omega})\big)\big(g(\hat{\boldsymbol{\omega}}^k) - g(\boldsymbol{\omega})\big),$$

它类似于离散情形 f 的梯度与 g 的梯度的点乘积. 最后, 对满足 $(\boldsymbol{\mu})_{\boldsymbol{\omega}} > 0$ $(\forall \boldsymbol{\omega} \in \mathbb{S})$ 的概率向量 $\boldsymbol{\mu}$, 定义

$$\mathcal{E}^{\boldsymbol{\mu}}(f,g) = \frac{1}{2}\langle \Gamma(f,g)\rangle_{\boldsymbol{\mu}}, \quad \rho_k^{\boldsymbol{\mu}}(\boldsymbol{\omega}) = \frac{(\boldsymbol{\mu})_{\boldsymbol{\omega}} + (\boldsymbol{\mu})_{\hat{\boldsymbol{\omega}}^k}}{2(\boldsymbol{\mu})_{\boldsymbol{\omega}}},$$

$$(\boldsymbol{Q}^{\boldsymbol{\mu}})_{\boldsymbol{\omega}\boldsymbol{\eta}} = \begin{cases} \rho_k^{\boldsymbol{\mu}}(\boldsymbol{\omega}), & \text{若 } \boldsymbol{\eta} = \hat{\boldsymbol{\omega}}^k; \\ 0, & \text{若 } \boldsymbol{\eta} \notin \{\boldsymbol{\omega}, \hat{\boldsymbol{\omega}}^1, \cdots, \hat{\boldsymbol{\omega}}^N\}; \\ -\sum\limits_{k=1}^{N} \rho_k^{\boldsymbol{\mu}}(\boldsymbol{\omega}), & \text{若 } \boldsymbol{\eta} = \boldsymbol{\omega}. \end{cases}$$

(a) 验证 $\boldsymbol{Q}^{\boldsymbol{\mu}}$ 是 $\mathbb{S}$ 上的不可约 Q-矩阵, 并满足逐一平衡条件 $(\boldsymbol{\mu})_{\boldsymbol{\omega}} \boldsymbol{Q}^{\boldsymbol{\mu}}_{\boldsymbol{\omega}\boldsymbol{\eta}} = (\boldsymbol{\mu})_{\boldsymbol{\eta}} \boldsymbol{Q}^{\boldsymbol{\mu}}_{\boldsymbol{\eta}\boldsymbol{\omega}}$. 进一步, 证明

$$-\langle g, \boldsymbol{Q}^{\boldsymbol{\mu}} f\rangle_{\boldsymbol{\mu}} = \mathcal{E}^{\boldsymbol{\mu}}(f,g).$$

(b) 令 $\boldsymbol{\lambda}$ 是定义在 $\mathbb{S}$ 上的均匀概率, 即 $(\boldsymbol{\lambda})_{\boldsymbol{\omega}} = 2^{-N}$, $\boldsymbol{\omega} \in \mathbb{S}$. 证明

$$\mathrm{Var}_{\boldsymbol{\mu}}(f) \leqslant M_{\boldsymbol{\mu}} \mathrm{Var}_{\boldsymbol{\lambda}}(f), \quad \text{其中 } M_{\boldsymbol{\mu}} = 2^N \max_{\boldsymbol{\omega}\in\mathbb{S}} (\boldsymbol{\mu})_{\boldsymbol{\omega}}$$

和

$$\mathcal{E}^{\boldsymbol{\lambda}}(f,f) \leqslant \frac{1}{m_{\boldsymbol{\mu}}}\mathcal{E}^{\boldsymbol{\mu}}(f,f), \quad \text{其中 } m_{\boldsymbol{\mu}} = 2^N \min_{\boldsymbol{\omega}\in\mathbb{S}}(\boldsymbol{\mu})_{\boldsymbol{\omega}}.$$

(c) 对每个 $S \subseteq \{1,\cdots,N\}$, 定义 $\chi_S : \mathbb{S} \to \{-1,1\}$, 使得 $\chi_S(\boldsymbol{\omega}) = \prod_{k\in S} \omega_k$. 特别地, $\chi_\varnothing = \mathbf{1}$. 证明 $\{\chi_S : S \subseteq \{1,\cdots,N\}\}$ 是 $L^2(\boldsymbol{\lambda})$ 上的一个正交基, 并且 $\boldsymbol{Q}^{\boldsymbol{\lambda}}\chi_S = -2(\#S)\chi_S$. 由此推出 $\mathrm{Var}_{\boldsymbol{\lambda}}(f) \leqslant \frac{1}{2}\mathcal{E}^{\boldsymbol{\lambda}}(f,f)$.

(d) 结合 (b) 和 (c), 证明 $\beta_{\boldsymbol{\mu}}\mathrm{Var}_{\boldsymbol{\mu}}(f) \leqslant \mathcal{E}^{\boldsymbol{\mu}}(f,f)$, 其中 $\beta_{\boldsymbol{\mu}} \equiv \frac{2m_{\boldsymbol{\mu}}}{M_{\boldsymbol{\mu}}}$. 特别地, 如果 $\{\boldsymbol{P}_t^{\boldsymbol{\mu}} : t \geqslant 0\}$ 是由 $\boldsymbol{Q}^{\boldsymbol{\mu}}$ 确定的转移概率矩阵半群, 证明 $\|\boldsymbol{P}_t^{\boldsymbol{\mu}} f - \langle f\rangle_{\boldsymbol{\mu}}\|_{2,\boldsymbol{\mu}} \leqslant e^{-\beta_{\boldsymbol{\mu}} t}\|f - \langle f\rangle_{\boldsymbol{\mu}}\|_{2,\boldsymbol{\mu}}$.

5.6.15. 回顾习题 5.6.14 中的 (b) 和 (c), 会有点意外地发现均匀概率 $\boldsymbol{\lambda}$ 的谱隙为 2, 不依赖于 N. 特别地, 这意味着如果 $\{\boldsymbol{P}^{(N)}(t) : t \geqslant 0\}$ 是 $\{-1,1\}^N$ 上由下述 Q-矩阵确定的半群:

$$(\boldsymbol{Q}^{(N)})_{\boldsymbol{\omega}\boldsymbol{\eta}} = \begin{cases} 1, & \text{若 } \boldsymbol{\eta} = \hat{\boldsymbol{\omega}}^k; \\ -N, & \text{若 } \boldsymbol{\eta} = \boldsymbol{\omega}; \\ 0, & \text{其他}. \end{cases}$$

则对任意的 $t \geqslant 0$ 和 $f \in L^2(\boldsymbol{\lambda}^{(N)})$(其中 $\boldsymbol{\lambda}^{(N)}$ 是 $\{-1,1\}^N$ 上的均匀概率测度), $\left\|\boldsymbol{P}^{(N)}(t)f - \langle f\rangle_{\boldsymbol{\lambda}^{(N)}}\right\|_{2,\boldsymbol{\lambda}^{(N)}} \leqslant e^{-2t}\|f\|_{2,\boldsymbol{\lambda}^{(N)}}$ 成立. 事实上, 这种情形有力地说明了这一章得到的理论在 Doeblin 理论注定行不通的一些情形也是有效的. 本题的目的在于证明, 对任意的 $t > 0$ 和 $\boldsymbol{\omega} \in \{-1,1\}^N$, 当 $\left(\boldsymbol{\mu}^{(N)}(t,\boldsymbol{\omega})\right)_{\boldsymbol{\eta}} = \left(\boldsymbol{P}^{(N)}(t)\right)_{\boldsymbol{\omega}\boldsymbol{\eta}}$ 时, $\lim_{N\to\infty}\left\|\boldsymbol{\mu}^{(N)}(t,\boldsymbol{\omega}) - \boldsymbol{\lambda}^{(N)}\right\|_v = 2$.

(a) 首先证明 $\left\|\boldsymbol{\mu}^{(N)}(t,\boldsymbol{\omega}) - \boldsymbol{\lambda}^{(N)}\right\|$ 不依赖于 $\boldsymbol{\omega}$.

(b) 证明对于可数空间 $\mathbb{S}$ 上的任意两个概率向量 $\boldsymbol{\nu}$ 和 $\boldsymbol{\nu}'$, 有 $2 \geqslant \|\boldsymbol{\nu} - \boldsymbol{\nu}'\|_v \geqslant 2|\boldsymbol{\nu}(A) - \boldsymbol{\nu}'(A)|$, $\forall A \subseteq \mathbb{S}$.

(c) 对 $1 \leqslant k \leqslant N$, 定义 $\{-1,1\}^N$ 上的随机变量 X_k 使得对 $\boldsymbol{\eta} = (\eta_1,\cdots,\eta_N)$, $X_k(\boldsymbol{\eta}) = \eta_k$. 证明在 $\boldsymbol{\lambda}^{(N)}$ 下, 这些 X_k 是相互独立的、$\{-1,1\}$ 值 Bernoulli 随机变量, 其期望为 0. 此外, 对给定的 $t > 0$, 记 $\boldsymbol{\mu}^{(N)} = \boldsymbol{\mu}^{(N)}(t,\boldsymbol{\omega})$, 其中 $\boldsymbol{\omega}$ 为 $\{-1,1\}^N$ 中的元素, 坐标分量都为 -1. 证明在 $\boldsymbol{\mu}^{(N)}$ 下, 这些 X_k 是相互独立的、$\{-1,1\}$ 值 Bernoulli 随机变量, 其期望值为 $-e^{-2t}$.

(d) 沿用 (c) 中的记号, 令 $A^{(N)}$ 为满足 $\frac{1}{N}\sum_{k=1}^{N} X_k(\boldsymbol{\eta}) \leqslant -\frac{1}{2}e^{-2t}$ 的 $\boldsymbol{\eta}$ 组成的集合, 证明 $\boldsymbol{\mu}^{(N)}(A^{(N)}) - \boldsymbol{\lambda}^{(N)}(A^{(N)}) \geqslant 1 - \frac{8e^{4t}}{N}$. 事实上, 利用 §1.2.4 末尾中推导出的那类估计, 特别是 (1.2.16), 要证明的不等式可以改进为 $\boldsymbol{\mu}^{(N)}(A^{(N)}) - \boldsymbol{\lambda}^{(N)}(A^{(N)}) \geqslant 1 - 2\exp\left(-\frac{Ne^{-4t}}{8}\right)$.

提示: 利用证明弱大数律时用过的 Chebyschev 估计.

(e) 综合上述得到, 对任意 $t > 0$, $N \in \mathbb{Z}^+$ 和 $\boldsymbol{\omega} \in \{-1, 1\}^N$, 有 $\left\|\boldsymbol{\mu}^{(N)}(t, \boldsymbol{\omega}) - \boldsymbol{\lambda}^{(N)}\right\|_v \geqslant 2\left(1 - \frac{8e^{4t}}{N}\right)$.

第六章 测度理论简介

1933 年的复活节那天, Kolmogorov 出版了他的《概率理论基础》一书, 此书奠定了绝大多数概率理论的基础. 由于 Kolmogorov 的模型是用 Lebesgue 的测度和积分理论的语言来叙述的, 因此, 若要对该模型有充分的认识, 必须首先对测度和积分理论有完整的理解. 虽然通过短短的一章无法使读者对该理论有深刻的认识, 我们还是想尽可能地利用有限的篇幅向读者介绍 Lebesgue 的基本思想以及 Kolmogorov 在他的概率论模型中对该理论的应用.

6.1 Lebesgue 测度理论

在这一节, 我们将介绍 Lebesgue 理论中的一些术语, 但是对于那些比较困难的结论的严格证明, 我们有意在此略去. 在其他很多书中可以找到这些证明, 比如 [8].

6.1.1. 测度空间: 测度理论中的基本元素包括以下三者: 一个集合 Ω (称为空间), Ω 的某些子集所组成的集类 $\mathcal{F}$ (称为*可测子集类*), 从 $\mathcal{F}$ 到 $[0,\infty]$ 的一个函数 μ (称为*测度*). $(\Omega,\mathcal{F})$ 被称作*可测空间*(在此空间上可定义测度). 而当 $(\Omega,\mathcal{F})$ 被赋予一个*测度* μ 时, $(\Omega,\mathcal{F},\mu)$ 被称为*测度空间*.

为了避免一些意义不大的平凡情况, 我们总假定空间 Ω 是非空的. 此外, 我们也假定可测子集类 $\mathcal{F}$ 构成 Ω 上的一个 σ-代数:

$$\Omega \in \mathcal{F}, \quad A \in \mathcal{F} \Longrightarrow A^c \equiv \Omega \setminus A \in \mathcal{F},$$
$$\text{且} \quad \{A_n\}_1^\infty \subseteq \mathcal{F} \Longrightarrow \bigcup_1^\infty A_n \in \mathcal{F}.$$

必须强调的是, 为了与点集拓扑区分开来 (即开集与闭集的描述), 在测度理论中, 只允许有限次或者可数次的集合运算.

利用初等的集合运算, 容易验证:

$$A, B \in \mathcal{F} \Longrightarrow A \bigcap B \in \mathcal{F}, \quad B \setminus A \in \mathcal{F},$$
$$\{A_n\}_1^\infty \subseteq \mathcal{F} \Longrightarrow \bigcap_1^\infty A_n \in \mathcal{F}.$$

最后, 测度 μ 必须把 $\varnothing$ 映成[1] 0, 且 μ 为**可数可加**, 即

$$\{A_n\}_1^\infty \subseteq \mathcal{F}, \quad \text{且当 } m \neq n \text{ 时}, \ A_m \bigcap A_n = \varnothing$$
$$\Longrightarrow \mu\left(\bigcup_1^\infty A_n\right) = \sum_1^\infty \mu(A_n). \tag{6.1.1}$$

特别地, 对于 $A, B \in \mathcal{F}$,

$$\begin{aligned} &A \subseteq B \Longrightarrow \mu(B) = \mu(A) + \mu(B \setminus A) \geqslant \mu(A), \\ &\mu(A \bigcap B) < \infty \Longrightarrow \mu(A \bigcup B) = \mu(A) + \mu(B) - \mu(A \bigcap B). \end{aligned} \tag{6.1.2}$$

上式第一行是把 B 写成两个互不相交集合 A 与 $B \setminus A$ 的并, 第二行是把 $A \bigcup B$ 写成两个互不相交集合 A 与 $B \setminus (A \bigcap B)$ 的并, 再把第一行的结论用于 B 和 $A \bigcap B$. 此外, 在 $\mu(B) = \mu(A \bigcap B) + \mu(B \setminus (A \bigcap B))$ 中, 当把 $\mu(A \bigcap B)$ 移到左边时, 必须有 "有限" 这个条件作保证, 也就是说, 要避免 ∞ 减 ∞.

当集合 Ω 为有限或可数时, 很容易构建测度空间. 具体来说, 可以令 $\mathcal{F} = \{A : A \subseteq \Omega\}$, 即 Ω 的子集全体. 作 Ω 的任意映射 $\mu : \omega \in \Omega \longmapsto \mu(\{\omega\}) \in [0, \infty]$, 由测度 μ 的可数可加性, 对于 $A \subseteq \Omega$, 必须有

$$\mu(A) = \sum_{\omega \in A} \mu(\{\omega\}).$$

但是, 当 Ω 为不可数时, 想要在非平凡可测子集类上构造感兴趣的测度并不是很容易的事. 实际上, 可以说, Lebesgue 的最大成就在于建

[1] 由可加性可得, 要么 $\mu(\varnothing) = 0$, 要么对一切 $A \in \mathcal{F}$, $\mu(A) = \infty$. 事实上, 由可加性, $\mu(\varnothing) = \mu(\varnothing \bigcup \varnothing) = 2\mu(\varnothing)$, 因此, $\mu(\varnothing)$ 要么为 0 要么为 ∞. 进一步, 若 $\mu(\varnothing) = \infty$, 则对任意的 $A \in \mathcal{F}$, $\mu(A) = \mu(A \bigcup \varnothing) = \mu(A) + \mu(\varnothing) = \infty$.

立了一个 $\Omega=\mathbb{R}$ 上的测度空间, 其中 $\mathcal{F}$ 是一个包括了所有开的、闭的、半开半闭的区间的 σ-代数, μ 把每个区间映成它的长度, 即若区间 I 的右端点和左端点分别为 b 和 a, 则 $\mu(I)=b-a$.

若 $\mu(\Omega)<\infty$, 称测度空间 $(\Omega,\mathcal{F},\mu)$ 为**有限的**. 这样的说法有些误导. 事实上, 此处的 "有限" 不是由 Ω 的大小决定的, 而是看 Ω 用 μ 来度量时到底有多大. 有些时候, 虽然一个测度空间并不是有限的, 但它可以分成可数个测度有限的集合, 这样的测度空间被称为是 σ-有限的. 等价地, 若存在 $\{\Omega_n\}_1^\infty\subseteq\mathcal{F}$ 使得[2] $\Omega_n\nearrow\Omega$, 且对于任意 $n\geqslant 1$, 有 $\mu(\Omega_n)<\infty$, 称 $(\Omega,\mathcal{F},\mu)$ 是 **σ-有限的**. 由此, $\mathbb{R}$ 上的 Lebesgue 测度和 $\mathbb{Z}$ 上的计数测度都是 σ-有限的, 但并不是有限的.

6.1.2. 关于可数可加性的一些结论: 可数可加性是该领域的一个核心. 特别地, 由此性质可导出下面的**测度连续性**:

$$\begin{aligned}&\{A_n\}_1^\infty\subseteq\mathcal{F}\ \text{和}A_n\nearrow A\Longrightarrow\mu(A_n)\nearrow\mu(A),\\&\{A_n\}_1^\infty\subseteq\mathcal{F},\mu(A_1)<\infty\ \text{和}A_n\searrow A\Longrightarrow\mu(A_n)\searrow\mu(A).\end{aligned}\tag{6.1.3}$$

虽然我们不打算证明本章中的多数结论, 但 (6.1.3) 式的证明简单且具有基本作用, 因此在下面给出. 事实上, 为了证明第一行, 只需令 $B_1=A_1$ 和 $B_{n+1}=A_{n+1}\setminus A_n$. 则当 $m\neq n$ 时, $B_m\bigcap B_n=\varnothing$, 对于任意 $n\geqslant 1$, $\bigcup\limits_1^n B_m=A_n$, 且 $\bigcup\limits_1^\infty B_m=A$. 因此, 由 (6.1.1) 式,

$$\mu(A_n)=\sum_1^n\mu(B_m)\nearrow\sum_1^\infty\mu(B_m)=\mu\left(\bigcup_1^\infty B_m\right)=\mu(A).$$

为了证明 (6.1.3) 式的第二行, 注意到 $\mu(A_1)=\mu(A_n)+\mu(A_1\setminus A_n)$ 和 $\mu(A_1)=\mu(A)+\mu(A_1\setminus A)$. 因为 $A_1\setminus A_n\nearrow A_1\setminus A$ 且 $\mu(A_1)<\infty$, 我们有 $\mu(A_1)-\mu(A_n)\nearrow\mu(A_1)-\mu(A)$, 因而 $\mu(A_n)\searrow\mu(A)$. 正如对 (6.1.2) 式中第二行的证明, 为了避免出现 ∞ 减 ∞ 的情形, 我们仍需要 "有限" 这个条件.

可数可加性的另一个重要推论是**次可数可加性**:

$$\{A_n\}_1^\infty\subseteq\mathcal{F}\Longrightarrow\mu\left(\bigcup_1^\infty A_n\right)\leqslant\sum_1^\infty\mu(A_n).\tag{6.1.4}$$

[2] 若对于一切 $n\geqslant 1$, $A_n\subseteq A_{n+1}$ 且 $A=\bigcup\limits_1^\infty A_n$, 我们记成 $A_n\nearrow A$. 类似地, $A_n\searrow A$ 表示对于一切 $n\geqslant 1$, $A_{n+1}\subseteq A_n$ 且 $A=\bigcap\limits_1^\infty A_n$. 显然 $A_n\nearrow A$ 当且仅当 $A_n^c\searrow A^c$.

类似前面的讨论, 上式的证明也是容易的. 具体来说, 令 $B_1 = A_1$ 和 $B_{n+1} = A_{n+1} \setminus \bigcup_1^n A_m$, 则 $\mu(B_n) \leqslant \mu(A_n)$ 且

$$\mu\left(\bigcup_1^\infty A_n\right) = \mu\left(\bigcup_1^\infty B_n\right) = \sum_1^\infty \mu(B_n) \leqslant \sum_1^\infty \mu(A_n).$$

(6.1.4) 式的一个特别重要的推论是

$$\text{若对于每一 } n \geqslant 1,\ \mu(A_n) = 0,\ \text{则 } \mu\left(\bigcup_1^\infty A_n\right) = 0. \tag{6.1.5}$$

也即测度为零的集合的可数并仍然是测度为零的集合. 下面我们将看到为什么在测度论中只考虑可数运算. 也就是说, 测度为零的集合的不可数并是否测度一定为零是不能肯定的. 例如, 考虑 $\mathbb{R}$ 上的 Lebesgue 测度, 由 (6.1.3) 式的第二行, 对任意的 $x \in \mathbb{R}$ 有

$$\mu\big(\{x\}\big) = \lim_{\delta \searrow 0} \mu\big((x-\delta, x+\delta)\big) = \lim_{\delta \searrow 0} 2\delta = 0.$$

但是, $(0,1) = \bigcup_{x \in (0,1)} \{x\}$ 的测度为 1.

6.1.3. 生成 σ-代数: 人们常常需要确定某一族子集是否在可测的子集类中. 为此, 了解下述构造很重要. 首先, 假设 $\mathcal{C}$ 是 Ω 的一个子集类. 然后存在一个 Ω 上包含 $\mathcal{C}$ 的最小 σ-代数, 称为*由 $\mathcal{C}$ 生成的 σ-代数*. 换言之, 考察 Ω 上包含 $\mathcal{C}$ 的所有 σ-代数的代数类. 这个代数类是非空的, 因为 $\{A : A \subseteq \Omega\}$ 是其中的元素. 另外, 容易验证任意一族 σ-代数的交仍然是 σ-代数. 从而, $\sigma(\mathcal{C})$ 是所有包含 $\mathcal{C}$ 的 σ-代数的交. 当 Ω 是一个拓扑空间并且 $\mathcal{C}$ 是 Ω 所有开子集组成的类时, $\sigma(\mathcal{C})$ 称为 *Borel σ-代数*, 记做 $\mathcal{B}_\Omega$.

要了解 σ-代数是如何生成的一个最重要的原因是, 人们为了验证 $\sigma(\mathcal{C})$ 中测度的性质, 常常只需验证这些性质在 $\mathcal{C}$ 上成立就够了. 下面的唯一性定理是一个重要的例子.

6.1.6 定理. *假设 $(\Omega, \mathcal{F})$ 为可测空间, 集类 $\mathcal{C} \subseteq \mathcal{F}$ 包含 Ω 并且对交运算是封闭的 (如果 $A, B \in \mathcal{C}$, 则 $A \bigcap B \in \mathcal{C}$). 若 μ 和 ν 是 $(\Omega, \mathcal{F})$ 上的一对有限测度, 并且对每个 $A \in \mathcal{C}$ 有 $\mu(A) = \nu(A)$, 则对所有的 $A \in \sigma(\mathcal{C})$ 有 $\mu(A) = \nu(A)$.*

证明: 我们称 $\mathcal{S} \subseteq \mathcal{F}$ 是 "好的", 如果

$$A, B \in \mathcal{S}, \quad A \subseteq B \Rightarrow B \setminus A \in \mathcal{S}, \tag{i}$$

$$A, B \in \mathcal{S}, \quad A \bigcap B = \varnothing \Rightarrow A \bigcup B \in \mathcal{S}, \tag{ii}$$

$$\{A_n\}_1^\infty \subseteq \mathcal{S}, \quad A_n \nearrow A \Rightarrow A \in \mathcal{S}. \tag{iii}$$

注意到如果 $\mathcal{S}$ 是 "好的", $\Omega \in \mathcal{S}$, 并且 $A, B \in \mathcal{S} \Rightarrow A \bigcap B \in \mathcal{S}$, 则 $\mathcal{S}$ 是 σ-代数. 事实上, 由 (i) 和 (iii), 我们只要验证对 $A, B \in \mathcal{S}$ 有 $A \bigcup B \in \mathcal{S}$. 然而, 因为 $A \bigcup B = \big(A \setminus (A \bigcap B)\big) \bigcup B$, 由 (i)、(ii) 和 $\mathcal{S}$ 对交运算封闭这一事实, 很容易得到上述结论. 另外, 注意到如果 $\mathcal{S}$ 是 "好的", 则对任意的 $\mathcal{D} \subseteq \mathcal{F}$,

$$\mathcal{S}' \equiv \{A \in \mathcal{S} : \text{对所有的 } B \in \mathcal{D}, \ A \bigcap B \in \mathcal{S}\}$$

也是 "好的".

现在记 $\mathcal{B} = \{A \in \mathcal{F} : \mu(A) = \nu(A)\}$. 由有限测度的性质, 特别是 (6.1.2) 和 (6.1.3) 式, 容易验证 $\mathcal{B}$ 是 "好的". 进一步, 由假设知 $\mathcal{C} \subseteq \mathcal{B}$. 从而, 若记 $\mathcal{B}' = \{A \in \mathcal{B} : \text{对所有的 } C \in \mathcal{C}, \ A \bigcap C \in \mathcal{B}\}$, 则由前面注意到的事实, $\mathcal{B}'$ 也是 "好的". 另外, 因为 $\mathcal{C}$ 对交运算封闭, 所以 $\mathcal{C} \subseteq \mathcal{B}'$. 类似地, $\mathcal{B}'' \equiv \{A \in \mathcal{B}' : \text{对所有的 } B \in \mathcal{B}', \ A \bigcap B \in \mathcal{B}'\}$ 也是 "好的", 并且, 由 $\mathcal{B}'$ 的定义, 知 $\mathcal{C} \subseteq \mathcal{B}''$. 最后, 如果 $A, A' \in \mathcal{B}''$ 且 $B \in \mathcal{B}'$, 则 $(A \bigcap A') \bigcap B = A \bigcap (A' \bigcap B) \in \mathcal{B}'$, 从而 $A \bigcap A' \in \mathcal{B}''$. 因此, $\mathcal{B}''$ 是包含 $\mathcal{C}$ 的 σ-代数, 且 $\mathcal{B}'' \subseteq \mathcal{B}$, 因而在 $\sigma(\mathcal{C})$ 上 μ 和 ν 相等. □

6.1.4. 可测函数: 给定一对可测空间 $(\Omega_1, \mathcal{F}_1)$ 和 $(\Omega_2, \mathcal{F}_2)$, 如果 $\mathcal{F}_2$ 中集合的逆像是 $\mathcal{F}_1$ 中的元素: 对每个 $\Gamma \in \mathcal{F}_2$ 有 $F^{-1}(\Gamma) \in \mathcal{F}_1$, 我们称映射 $F : \Omega_1 \to \Omega_2$ 是**可测的**.[3] 注意到当 $\mathcal{F}_2 = \sigma(\mathcal{C})$ 时, 若对每个 $C \in \mathcal{C}$, $F^{-1}(C) \in \mathcal{F}_1$, 则 F 是可测的. 特别地, 若 Ω_1 和 Ω_2 是拓扑空间, $\mathcal{F}_i = \mathcal{B}_{\Omega_i}$, 则每一个从 Ω_1 到 Ω_2 的连续映射是可测的.

了解下列结论很重要. 即当 $\Omega_2 = \mathbb{R}$ 和 $\mathcal{F} = \mathcal{B}_{\mathbb{R}}$ 时, **可测性对序列极限运算保持不变**. 确切地说, 若 $\{f_n\}_1^\infty$ 是从 $(\Omega, \mathcal{F})$ 到 $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ 的 $\mathbb{R}$ 值可测函数序列, 则

$$\begin{aligned} &\omega \rightsquigarrow \sup_n f_n(\omega), \qquad \omega \rightsquigarrow \inf_n f_n(\omega), \\ &\omega \rightsquigarrow \limsup_{n\to\infty} f_n(\omega) \quad \text{和} \quad \omega \rightsquigarrow \liminf_{n\to\infty} f_n(\omega) \end{aligned} \quad \text{都是可测函数}. \tag{6.1.7}$$

[3] 读者可能已经注意到这一定义与通过开集之逆像定义连续性有惊人相似之处.

例如, 上面结论中的第一个可由下述推导证得. 首先注意到当 $\mathcal{C} = \{(a,\infty) : a \in \mathbb{R}\}$ 时, $\mathcal{B}_{\mathbb{R}} = \sigma(\mathcal{C})$. 因此, $f : \Omega \to \mathbb{R}$ 是可测的当且仅当对每个 $a \in \mathbb{R}$, $\{f > a\}$[4] 是可测的, 从而 (6.1.7) 式中的第一个函数是可测的, 这是因为

$$\left\{\sup_n f_n > a\right\} = \bigcup_n \{f_n > a\}, \quad \forall\, a \in \mathbb{R}.$$

对于 (6.1.7) 式第二行的推论, 我们知道使得 $\lim\limits_{n\to\infty} f_n(\omega)$ 存在的点 ω 组成的集合 Δ 是可测集, 并且若 $\Delta = \Omega$, 则 $\omega \rightsquigarrow \lim\limits_{n\to\infty} f_n(\omega)$ 是可测的.

最后, 可测函数引出前述小节中 σ-代数构造的一些重要例子. 那就是, 假设空间 Ω_1 和可测空间 $(\Omega_2, \mathcal{F}_2)$ 给定, 令 $\mathfrak{F}$ 为某些从 Ω_1 到 Ω_2 的映射组成的类. 则在 Ω_1 上由 $\mathfrak{F}$ 生成的 σ-代数 $\sigma(\mathfrak{F})$是 Ω_1 上的使得 $\mathfrak{F}$ 中每一个元素可测的最小 σ-代数. 等价地, $\sigma(\mathfrak{F}) = \sigma(\mathcal{C})$, 其中 $\mathcal{C} = \{F^{-1}(\Gamma) : F \in \mathfrak{F}, \Gamma \in \mathcal{F}_2\}$. 最后, 假设 $\mathcal{F}_2 = \sigma(\mathcal{C}_2)$, 其中 $\mathcal{C}_2 \subseteq \mathcal{F}_2$ 包含 Ω_2, 并对交运算是封闭的 (即若 $A, B \in \mathcal{C}_2$, 则 $A \bigcap B \in \mathcal{C}_2$), 再令 $\mathcal{C}_1$ 是形如 $F_1^{-1}(A_1) \bigcap \cdots \bigcap F_n^{-1}(A_n)$, $n \in \mathbb{Z}^+$ 的集合组成的类, 其中 $\{F_1, \cdots, F_n\} \subseteq \mathfrak{F}$, $\{A_1, \cdots, A_n\} \subseteq C_2$. 则 $\sigma(\mathfrak{F}) = \sigma(\mathcal{C}_1)$, $\Omega_1 \in \mathcal{C}_1$, 并且 $\mathcal{C}_1$ 对交运算封闭. 特别地, 根据定理 6.1.6, 如果 μ 和 ν 是 $(\Omega_1, \mathcal{F}_1)$ 上的一对有限测度, 并且

$$\begin{aligned}&\mu\big(\{\omega_1 : F_1(\omega_1) \in A_1, \cdots, F_n(\omega_1) \in A_n\}\big)\\ =\ &\nu\big(\{\omega_1 : F_1(\omega_1) \in A_1, \cdots, F_n(\omega_1) \in A_n\}\big)\end{aligned}$$

对所有 $n \in \mathbb{Z}^+$, $\{F_1, \cdots, F_n\} \subseteq \mathfrak{F}$ 和 $\{A_1, \cdots, A_n\} \subseteq C_2$ 成立, 则在 $\sigma(\mathfrak{F})$ 上 μ 和 ν 相等.

6.1.5. Lebesgue 积分: 给定测度空间 $(\Omega, \mathcal{F}, \mu)$, Lebesgue 的积分理论开始于给出一个定义所有非负可测函数 (即所有从 $(\Omega, \mathcal{F})$ 映到 $([0,\infty]; \mathcal{B}_{[0,\infty]})$[5] 的可测映射 f) 关于 μ 的积分的准则. 也就是说, 按他的理论, 当 $\mathbf{1}_A$ 为集合 $A \in \mathcal{F}$ 的示性函数且 $a \in [0,\infty)$ 时, 函数 $a\mathbf{1}_A$[6] 的积分应等于 $a\mu(A)$. 然后他强调积分应该有可加性, 即 $f_1 + f_2$ 的积分应为 f_1 和 f_2 的积分之和. 特别地, 这意味着, 如果 f 是非负可测函

[4] 在不至于产生混淆时, 我们用 $\{F \in \Gamma\}$ 表示 $\{\omega : F(\omega) \in \Gamma\}$.

[5] 在上下文中, 我们把 $[0,\infty]$ 看做是通过映射 $t \in [0,1] \longmapsto \tan\left(\frac{\pi}{2}t\right)$ 将 $[0,1]$ 映成 $[0,\infty]$ 而得到的紧致距离空间.

[6] 在测度论中, 习惯于取 $0 \cdot \infty = 0$, 这样最方便.

数, 并且是简单的, 即只取有限个值, 那么 f 的积分 $\int f d\mu$ 必须是

$$\sum_{x\in[0,\infty)} x\mu(f^{-1}(\{x\})),$$

其中, 由于除了有限个 x 外, 都有 $f^{-1}(\{x\}) = \varnothing$, 上述和式中只包含有限个非零的项. 当然, 在强调积分的这一可加性之前, 不得不验证可加性是一致的. 确切地说, 需要证明的是, 如果 $\sum_1^n a_m \mathbf{1}_{A_m} = f = \sum_1^{n'} a'_m \mathbf{1}_{A'_m}$, 则有 $\sum_1^n a_m\mu(A_m) = \sum_1^{n'} a'_m\mu(A'_m)$. 然而, 一旦建立了一致性, 人们就知道如何对任何非负可测简单函数定义积分, 使得积分是可加的, 并且对示性函数有显然的答案. 特别地, 可加性可推出**单调性**: 如果 $f \leqslant g$, 则 $\int f d\mu \leqslant \int g d\mu$.

为了完成 Lebesgue 关于非负函数的讨论, 首先必须注意到, 如果 $f : E \to [0,\infty]$ 是可测的, 那么存在一列非负可测简单函数 $\{\varphi_n\}_1^\infty$, 使得对每个 $\omega \in \Omega, \varphi_n(\omega) \nearrow f(\omega)$. 例如, 我们可取

$$\varphi_n = \sum_{m=1}^{4^n} \frac{m}{2^n} \mathbf{1}_{A_{m,n}}, \quad 其中\ A_{m,n} = \{\omega : m2^{-n} \leqslant f(\omega) < (m+1)2^{-n}\}.$$

给定 $\{\varphi_n\}_1^\infty$, Lebesgue 所说的是 $\int f d\mu = \lim_{n\to\infty} \int \varphi_n d\mu$. 事实上, 由单调性, $\int \varphi_n d\mu$ 关于 n 非降, 从而提到的极限一定存在. 另一方面, 和前面提到的一样, 在采纳 Lebesgue 的定义之前, 一致性的问题必须解决. 即有千万种方法去构造用来逼近的简单函数 φ_n, 我们必须保证上述极限不依赖于人们所选取的逼近方案. 就是说, 必须验证, 如果 $\{\varphi_n\}_1^\infty$ 和 $\{\psi_n\}_1^\infty$ 为两列非负简单可测函数, 使得对每一 $\omega \in \Omega$, $\lim_{n\to\infty} \varphi_n(\omega) = \lim_{n\to\infty} \psi_n(\omega)$, 则 $\lim_{n\to\infty} \int \varphi_n(\omega) d\mu = \lim_{n\to\infty} \int \psi_n(\omega) d\mu$; 正是在这一步, 必须充分利用可数可加性.

对所有非负可测函数 f 定义了积分 $\int f d\mu$ 后, 必须验证所定义的积分具有齐次性和可加性: 对 $a \in [0,\infty]$, $\int af d\mu = a\int f d\mu$ 且 $\int (f_1 + f_2) d\mu = \int f_1 d\mu + \int f_2 d\mu$. 然而, 这些性质容易由简单函数 f 的情形得到. 因而, 在 Lebesgue 积分的构造中, 剩下的是要去掉非负函数的限制, 将积分推广到符号可正可负的可测函数. 另一方面, 如果想保持积分的线性, 那么如何做这一推广是毫无疑问的了. 也即, 给定一个可测函数 $f : E \to [-\infty,\infty]$, 不难证明 $f^+ \equiv f \vee 0$ 和 $f^- \equiv -(f \wedge 0) = (-f)^+$

都是非负可测函数. 从而, 由于 $f = f^+ - f^-$, 线性要求 $\int f d\mu = \int f^+ d\mu - \int f^- d\mu$; 这时必须面对两个问题. 首先, 考察范围必须限制到 $\int f^+ d\mu$ 和 $\int f^- d\mu$ 至少有一个是有限的函数 f, 否则会出现致命的 $\infty - \infty$ 的情形. 其次, 必须再次验证一致性, 即如果 $f = f_1 - f_2$, 其中 f_1 和 f_2 为非负可测函数, 那么 $\int f_1 d\mu - \int f_2 d\mu = \int f^+ d\mu - \int f^- d\mu$.

在大多数应用中, 求积分的可测函数或者是非负的或者满足 $\int |f| d\mu < \infty$, 具有该性质的函数 f 称为可积函数. 因为 $|a_1 f_1 + a_2 f_2| \leqslant |a_1||f_1| + |a_2||f_2|$, 可积函数组成的类构成 $\mathbb{R}$ 上的一个向量空间, 在这个向量空间上关于 μ 的积分运算相当于一个线性泛函. 最后, 如果 f 为可测函数, 且它或者非负或者可积, $A \in \mathcal{F}$, 则乘积 $\mathbf{1}_A f$ 也是可测函数, 或者非负或者可积, 从而

$$\int_A f d\mu \equiv \int \mathbf{1}_A f d\mu \tag{6.1.8}$$

有定义.

6.1.6. Lebesgue 积分的稳定性: Lebesgue 理论的有效性来源于所定义的积分的稳定性, 它的稳定性通过下面三个著名的定理体现出来. 以下所有可测性和可积性都是相对于测度空间 $(\Omega, \mathcal{F}, \mu)$ 而言的. 并且, 假定所有函数的值域是 $(-\infty, \infty]$.

6.1.9 定理. (单调收敛定理) *设 $\{f_n\}_1^\infty$ 是一列非降可测函数, 对每个 $\omega \in \Omega$, 令 $f(\omega) = \lim_{n\to\infty} f_n(\omega)$. 如果存在一个被所有 f_n 控制的可积函数 g, 则 $\int f_n d\mu \nearrow \int f d\mu$. 相应地, 如果$\{f_n\}_1^\infty$ 是非增的, 且存在一个控制每个 f_n 的可积函数 g, 则 $\int f_n d\mu \searrow \int f d\mu$.*

6.1.10 定理. (Fatou 引理) *给定任意可测函数列 $\{f_n\}_1^\infty$, 每个 f_n 控制可积函数 g, 则*

$$\liminf_{n\to\infty} \int f_n d\mu \geqslant \int \liminf_{n\to\infty} f_n d\mu.$$

相应地, 若存在某个控制所有 f_n 的可积函数 g, 则

$$\limsup_{n\to\infty} \int f_n d\mu \leqslant \int \limsup_{n\to\infty} f_n d\mu.$$

6.1.11 定理. (Lebesgue 控制收敛定理) *设 $\{f_n\}_1^\infty$ 是可测函数序列. 假定存在一个可积函数 g 使得对所有 $n \geqslant 1$, $\mu(\{\omega : |f_n(\omega)| > g(\omega)\}) = 0$. 并且假定 $\{f_n\}_1^\infty$ 以下面两种方式之一收敛到可测函数 f:*

(a) $\mu\left(\left\{\omega : f(\omega) \neq \lim_{n\to\infty} f_n(\omega)\right\}\right) = 0$;

(b) $\lim_{n\to\infty} \mu(\{\omega : |f_n(\omega) - f(\omega)| \geqslant \epsilon\}) = 0$, 任意 $\epsilon > 0$.

则当 $n \to \infty$ 时,

$$\left|\int f_n d\mu - \int f d\mu\right| \leqslant \int |f_n - f| d\mu \to 0.$$

在继续讨论之前, 我们要对 Lebesgue 控制收敛定理的条件作一个说明. 因为在一个**零测度集** (即满足 $\mu(A) = 0$ 的集合 $A \in \mathcal{F}$) 上, 我们无法看出积分会产生什么情况, 所以很自然地, 那些保证积分具有某种性质的条件应当只是在一个零测度集的补集上成立. 因此, 条件 (a) 是说对于零测度集外的所有 ω 有 $f_n(\omega) \to f(\omega)$. 用专业术语来说, 在一个零测度集以外成立的条件被称为**几乎处处成立的条件**. 根据这个术语, 第一个假设就是几乎处处有 $|f_n| \leqslant g$, 而 (a) 表示 $\{f_n\}_1^\infty$ 几乎处处收敛到 f, 文献中常简写为 $|f_n| \leqslant g$ a.e. 和 $f_n \to f$ a.e.. 条件 (b) 和 (a) 有关, 但有很大差别. 特别地, (b) 并不保证对于所有的 $\omega \in \Omega$, $\{f_n(\omega)\}_1^\infty$ 都是收敛的. 例如, 令 μ 是上面所定义的 $\mathbb{R}$ 上的 Lebesgue 测度, $f_{m+2^n}(\omega) = \mathbf{1}_{[0,2^{-n})}(\omega - m2^{-n})$, $n \geqslant 0$, $0 \leqslant m < 2^n$. 则 $\mu(\{\omega : f_{m+2^n}(\omega) \neq 0\}) = 2^{-n}$. 因此对于任意的 $\epsilon > 0$, $\lim_{n\to\infty} \mu(\{\omega : |f_n(\omega)| \geqslant \epsilon\}) = 0$. 另一方面, 对于每个 $\omega \in [0,1)$, $\limsup_{n\to\infty} f_n(\omega) = 1$, 但 $\liminf_{n\to\infty} f_n(\omega) = 0$. 因此, 由 (b) 推不出 (a). 反之, 当 $\mu(\Omega) < \infty$ 时, 由 (a) 可以推出 (b), 但是当 $\mu(\Omega) = \infty$ 时, 由 (a) 推不出 (b). 为了明白这一点, 再次令 μ 是 Lebesgue 测度, 考虑 $f_n = \mathbf{1}_{\mathbb{R}\setminus[-n,n]}$.

与前面的讨论有关的是一个被称作 *Markov* 不等式的基本估计, 它在测度论分析中起着重要作用. 因为对任意 $\lambda > 0$, $\lambda\mathbf{1}_{[\lambda,\infty]} \circ f \leqslant f\mathbf{1}_{[\lambda,\infty]} \circ f \leqslant |f|$, 所以

$$\mu(\{\omega : f(\omega) \geqslant \lambda\}) \leqslant \frac{1}{\lambda}\int_{\{\omega:f(\omega)\geqslant\lambda\}} f d\mu \leqslant \frac{1}{\lambda}\int |f| d\mu. \tag{6.1.12}$$

由此可推出以下结论: 对于所有的 $\epsilon > 0$,

$$\lim_{n\to\infty}\int |f_n - f| d\mu = 0 \Longrightarrow \mu(\{\omega : |f_n(\omega) - f(\omega)| \geqslant \epsilon\}) = 0.$$

也即, 条件 (b) 对于 Lebesgue 定理中的结论是必要的. 另外, (6.1.12)

式证明了对所有的 $\epsilon > 0$,

$$\int |f| d\mu = 0 \Longrightarrow \mu(\{\omega : |f(\omega)| \geqslant \epsilon\}) = 0,$$

因此, 由 (6.1.3) 式,

$$\int |f| d\mu = 0 \Longrightarrow \mu(\{\omega : f(\omega) \neq 0\}) = 0. \tag{6.1.13}$$

最后, 通过考虑序列 $\{f_n\}_1^\infty$, 其中 $f_n = n\mathbf{1}_{[0,\frac{1}{n})}$ 或 $f_n = \mathbf{1}_{[n-1,n)}$ 且 μ 是 Lebesgue 测度, 就可以看清 Lebesgue 控制函数 g 的作用.

6.1.7. 可数空间上的 Lebesgue 积分: 在这一节中我们将讨论下列相对平凡的情形: Ω 是可数的, $\mathcal{F}$ 包含了 Ω 的所有子集且 μ 是 σ-有限的. 考察这时有怎样的 Lebesgue 理论. 如 §6.1.1 中所述, 在 $(\Omega, \mathcal{F})$ 上定义一个测度等价于给 Ω 的每个元素指定一个非负的数. 而且因为我们假定测度是 σ-有限的, 所以它不应为 ∞.

因为 Ω 的所有元素是可以被计数的, 所以我们没有理由不去计算它们的个数. 因此, 当 Ω 是有限时, 不失一般性, 我们记 $\Omega = \{1, \cdots, N\}$, 其中 $N = \#\Omega$ 表示 Ω 的元素个数, 类似地, 当 Ω 是可数无限时, 我们也可以 (至少是抽象地) 认为它是正整数集 $\mathbb{Z}^+$. 事实上, 为了避免将有限情形和可数无限情形分开讨论, 我们可以将有限情形嵌入到可数无限情形进行研究. 注意若对于 $\omega > N$, 我们假定 $\mu(\{\omega\}) = 0$, 那么 $\{1, \cdots, N\}$ 和 $\mathbb{Z}^+$ 两者是完全一致的. 最后, 为了上下文中符号的统一, 我们用符号 $\mathbb{S}$ 代替 Ω, 而 i, j, 或者 k 表示 $\mathbb{S}$ 的普通元素, 且相应于 μ 我们通过 $(\boldsymbol{\mu})_i = \mu(\{i\})$ 定义一个行向量测度 $\boldsymbol{\mu} \in [0, \infty)^{\mathbb{S}}$.

首先注意到, 当 $\int f^+ d\mu < \infty$ 或 $\int f^- d\mu < \infty$ 时,

$$\int f d\mu = \sum_{i \in \mathbb{S}} f(i)(\boldsymbol{\mu})_i. \tag{6.1.14}$$

事实上, 如果 $f \geqslant 0$ 是简单函数, f 可取有限个互不相同的值 $a_1, \cdots, a_L$, 且记 $A_\ell = \{i : f(i) = a_\ell\}$, 则

$$\begin{aligned}\int f d\mu &= \sum_{\ell=1}^{L} a_\ell \mu(A_\ell) = \sum_{\ell=1}^{L} a_\ell \sum_{i \in A_\ell} (\boldsymbol{\mu})_i \\ &= \sum_{\ell=1}^{L} \sum_{i \in A_\ell} f(i)(\boldsymbol{\mu})_i = \sum_{i \in \mathbb{S}} f(i)(\boldsymbol{\mu})_i.\end{aligned}$$

其次, 在 $\mathbb{S}$ 上给定任意的 $f \geqslant 0$, 当 $1 \leqslant i \leqslant n$ 时, 令 $\varphi_n(i) = f(i)$, 当 $i > n$ 时, 令 $\varphi_n(i) = 0$. 则

$$\int f d\mu = \lim_{n\to\infty} \int \varphi_n d\mu = \lim_{n\to\infty} \sum_{1\leqslant i\leqslant n} f(i)(\boldsymbol{\mu})_i = \sum_{i\in\mathbb{S}} f(i)(\boldsymbol{\mu})_i.$$

最后, 若 $\int f^+ d\mu$ 或 $\int f^- d\mu$ 是有限的, 则显然

$$\begin{aligned}\int f d\mu &= \int f^+ d\mu - \int f^- d\mu \\ &= \sum_{\{i:f(i)\geqslant 0\}} f(i)(\boldsymbol{\mu})_i - \sum_{\{i:f(i)\leqslant 0\}} f(i)(\boldsymbol{\mu})_i = \sum_{i\in\mathbb{S}} f(i)(\boldsymbol{\mu})_i.\end{aligned}$$

我们下面来看 "三大定理" 在这里起什么作用.

单调收敛定理: 首先指出我们只需要讨论 $0 \leqslant f_n \nearrow f$ 这种情形. 事实上, 我们只要用 $f_n - g$ 或 $g - f_n$ 代替 f_n 即可将定理中的两种情形简化为上述情形. 当 $0 \leqslant f_n \nearrow f$ 时, 显然

$$0 \leqslant \sum_{i\in\mathbb{S}} f_n(i)(\boldsymbol{\mu})_i \leqslant \sum_{i\in\mathbb{S}} f_{n+1}(i)(\boldsymbol{\mu})_i \leqslant \sum_{i\in\mathbb{S}} f(i)(\boldsymbol{\mu})_i.$$

余下只需注意到

$$\lim_{n\to\infty} \sum_{i\in\mathbb{S}} f_n(i)(\boldsymbol{\mu})_i \geqslant \lim_{n\to\infty} \sum_{i=1}^{L} f_n(i)(\boldsymbol{\mu})_i = \sum_{i=1}^{L} f(i)(\boldsymbol{\mu})_i,$$

其中 $L \in \mathbb{Z}^+$. 令 $L \nearrow \infty$ 即得所需的结论.

Fatou **引理**: 我们仍然可以简化到 $f_n \geqslant 0$ 的情形且限于讨论下极限. 这时候

$$\begin{aligned}\liminf_{n\to\infty} \sum_{i\in\mathbb{S}} f_n(i)(\boldsymbol{\mu})_i &= \lim_{m\to\infty} \inf_{n\geqslant m} \sum_{i\in\mathbb{S}} f_n(i)(\boldsymbol{\mu})_i \\ &\geqslant \lim_{m\to\infty} \sum_{i\in\mathbb{S}} \inf_{n\geqslant m} f_n(i)(\boldsymbol{\mu})_i = \sum_{i\in\mathbb{S}} \liminf_{n\to\infty} f_n(i)(\boldsymbol{\mu})_i,\end{aligned}$$

其中最后一步是根据单调收敛定理有 $0 \leqslant \inf_{n\geqslant m} f_n \nearrow \liminf_{n\to\infty} f_n$ 而推得的.

Lebesgue **控制收敛定理**: 首先注意到如果我们将那些满足 $(\boldsymbol{\mu})_i = 0$ 的元素 $i \in \mathbb{S}$ 去掉, 所有的结论都不会发生变化. 因此从现在开始我们假定对所有的 $i \in \mathbb{S}$ 有 $(\boldsymbol{\mu})_i > 0$. 其次我们指出, 在上述假定下, 定

理的假设条件变为对于每个 $i \in \mathbb{S}$ 有 $\sup_n |f_n(i)| \leqslant g(i)$ 和 $f_n(i) \to f(i)$. 特别地, $|f| \leqslant g$. 因此, 通过用 $f_n - f$ 代替 f_n, $2g$ 代替 g, 我们可以假定 $f \equiv 0$. 给定 $\epsilon > 0$, 选取 L 使得 $\sum_{i>L} g(i)(\boldsymbol{\mu})_i < \epsilon$. 则当 $n \to \infty$ 时,

$$\left|\sum_{i \in \mathbb{S}} f_n(i)(\boldsymbol{\mu})_i\right| \leqslant \sum_{i \in \mathbb{S}} |f_n(i)|(\boldsymbol{\mu})_i \leqslant \sum_{i=1}^{L} |f_n(i)|(\boldsymbol{\mu})_i + \epsilon \to \epsilon.$$

6.1.8. Fubini 定理: Fubini 定理研究的是测度空间的乘积. 即给定测度空间 $(\Omega_1, \mathcal{F}_1)$ 和 $(\Omega_2, \mathcal{F}_2)$, $\mathcal{F}_1$ 和 $\mathcal{F}_2$ 的乘积是 $\Omega_1 \times \Omega_2$ 上的 σ-代数 $\mathcal{F}_1 \times \mathcal{F}_2$, 其中 $\Omega_1 \times \Omega_2$ 是由可测矩形集 $\{A_1 \times A_2 : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}$ 构成的. 有关这个构造的一个重要性质是, 如果 f 是 $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2)$ 上的可测函数, 那么, 对每个 $\omega_1 \in \Omega_1$ 和 $\omega_2 \in \Omega_2$, $\omega_2 \rightsquigarrow f(\omega_1, \omega_2)$ 和 $\omega_1 \rightsquigarrow f(\omega_1, \omega_2)$ 是相应于 $(\Omega_2, \mathcal{F}_2)$ 和 $(\Omega_1, \mathcal{F}_1)$ 上的可测函数.

在下文中, 确切地强调哪个是被积变量是重要的. 基于这个原因, 我们使用更详细的记号 $\int_\Omega f(\omega)\mu(d\omega)$ 代替简写式 $\int f d\mu$.

6.1.15 定理. (Fubini 定理)[7] 设 $(\Omega_1, \mathcal{F}_1, \mu_1)$ 和 $(\Omega_2, \mathcal{F}_2, \mu_2)$ 是一对 σ-有限测度空间, 令 $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$. 则在 $(\Omega, \mathcal{F})$ 上存在一个唯一的测度 $\mu = \mu_1 \times \mu_2$, 使得对所有的 $A_1 \in \mathcal{F}_1$ 和 $A_2 \in \mathcal{F}_2$ 有 $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$. 并且, 如果 f 是 $(\Omega, \mathcal{F})$ 上的非负可测函数, 则

$$\omega_1 \rightsquigarrow \int_{\Omega_2} f(\omega_1, \omega_2)\mu_2(d\omega_2) \quad 和 \quad \omega_2 \rightsquigarrow \int_{\Omega_1} f(\omega_1, \omega_2)\mu_1(d\omega_1)$$

都是可测函数, 且

$$\begin{aligned}
&\int_{\Omega_1}\left(\int_{\Omega_2} f(\omega_1, \omega_2)\mu_2(d\omega_2)\right)\mu_1(d\omega_1) \\
&= \int_\Omega f\, d\mu = \int_{\Omega_2}\left(\int_{\Omega_1} f(\omega_1, \omega_2)\mu_1(d\omega_1)\right)\mu_2(d\omega_2).
\end{aligned}$$

[7] 尽管 Tonelli 对 Fubini 定理做出了很大的贡献, 但是很少被提到, 这个定理常常被归功于 Fubini.

最后, 对于 $(\Omega, \mathcal{F}, \mu)$ 上的任一可积函数 f,

$$A_1 = \left\{\omega_1 \ : \ \int_{\Omega_2} |\ f(\omega_1, \omega_2)\ |\ \mu_2(d\omega_2) < \infty\right\} \in \mathcal{F}_1,$$

$$A_2 = \left\{\omega_2 \ : \ \int_{\Omega_1} |\ f(\omega_1, \omega_2)\ |\ \mu_1(d\omega_1) < \infty\right\} \in \mathcal{F}_2,$$

$$\omega_1 \rightsquigarrow f_1(\omega_1) \equiv \mathbf{1}_{A_1}(\omega_1) \int_{\Omega_2} f(\omega_1, \omega_2)\mu_2(d\omega_2),$$

$$\omega_2 \rightsquigarrow f_2(\omega_2) \equiv \mathbf{1}_{A_2}(\omega_2) \int_{\Omega_1} f(\omega_1, \omega_2)\mu_1(d\omega_1)$$

都是可积的, 且

$$\int_{\Omega_1} f_1(\omega_1)\mu_1(d\omega_1) = \int_{\Omega} f(\omega)\mu(d\omega) = \int_{\Omega_2} f_2(\omega_2)\mu_2(d\omega_2).$$

当 Ω_1 和 Ω_2 是可数时, 一切就变得简单了. 我们用 §6.1.6 中的记号, 乘积测度 $\mu_1 \times \mu_2$ 相应于行向量 $\boldsymbol{\mu}_1 \times \boldsymbol{\mu}_2 \in [0, \infty)^{\mathbb{S}_1 \times \mathbb{S}_2}$, 其中 $(\boldsymbol{\mu}_1 \times \boldsymbol{\mu}_2)_{(i_1, i_2)} = (\boldsymbol{\mu})_{i_1}(\boldsymbol{\mu})_{i_2}$. 所以由 (6.1.14) 式, 当 $\{a_{i_1 i_2} : (i_1, i_2) \in \mathbb{S}_1 \times \mathbb{S}_2\} \subseteq (-\infty, \infty]$ 满足对所有的 (i_1, i_2), $a_{i_1 i_2} \geqslant 0$ 或者 $\sum\limits_{(i_1, i_2)} |a_{i_1 i_2}| < \infty$ 时, 由 Fubini 定理可得

$$\sum_{i_1 \in \mathbb{S}_1} \left(\sum_{i_2 \in \mathbb{S}_2} a_{i_1 i_2}\right) = \sum_{(i_1, i_2) \in \mathbb{S}_1 \times \mathbb{S}_2} a_{i_1 i_2} = \sum_{i_2 \in \mathbb{S}_2} \left(\sum_{i_1 \in \mathbb{S}_1} a_{i_1 i_2}\right).$$

在证明这个结论的过程中, 我们假定 $\mathbb{S}_1 = \mathbb{Z}^+ = \mathbb{S}_2$. 首先考虑 $a_{i_1 i_2} \geqslant 0$ 的情形. 任给 $(n_1, n_2) \in (\mathbb{Z}^+)^2$,

$$\sum_{(i_1, i_2) \in \mathbb{S}_1 \times \mathbb{S}_2} a_{i_1 i_2} \geqslant \sum_{\substack{(i_1, i_2) \in \mathbb{S}_1 \times \mathbb{S}_2 \\ i_1 \leqslant n_1, \ i_2 \leqslant n_2}} a_{i_1 i_2} = \sum_{i_2 = 1}^{n_2} \left(\sum_{i_1 = 1}^{n_1} a_{i_1 i_2}\right).$$

因此, 先令 $n_1 \to \infty$, 再令 $n_2 \to \infty$, 我们可以得到

$$\sum_{(i_1, i_2) \in \mathbb{S}_1 \times \mathbb{S}_2} a_{i_1 i_2} \geqslant \sum_{i_2 \in \mathbb{S}_2} \left(\sum_{i_1 \in \mathbb{S}_1} a_{i_1 i_2}\right).$$

同理, 任给 $n \in \mathbb{Z}^+$,

$$\sum_{\substack{(i_1, i_2) \in \mathbb{S}_1 \times \mathbb{S}_2 \\ i_1 \vee i_2 \leqslant n}} a_{i_1 i_2} = \sum_{i_2 = 1}^{n} \left(\sum_{i_1 = 1}^{n} a_{i_1 i_2}\right) \leqslant \sum_{i_2 \in \mathbb{S}_2} \left(\sum_{i_1 \in \mathbb{S}_1} a_{i_1 i_2}\right),$$

再令 $n \to \infty$, 我们得到反向的不等式. 而对 $\sum_{(i_1,i_2)} |a_{i_1 i_2}| < \infty$ 的情形, 对所有 $i_2 \in \mathbb{S}_2$,

$$\sum_{i_1 \in \mathbb{S}_1} |a_{i_1 i_2}| < \infty,$$

并因此

$$\begin{aligned}\sum_{(i_1,i_2)\in\mathbb{S}_1\times\mathbb{S}_2} a_{i_1 i_2} &= \sum_{(i_1,i_2)\in\mathbb{S}_1\times\mathbb{S}_2} a^+_{i_1 i_2} - \sum_{(i_1,i_2)\in\mathbb{S}_1\times\mathbb{S}_2} a^-_{i_1 i_2} \\ &= \sum_{i_2\in\mathbb{S}_2}\left(\sum_{i_1\in\mathbb{S}_1} a^+_{i_1 i_2}\right) - \sum_{i_2\in\mathbb{S}_2}\left(\sum_{i_1\in\mathbb{S}_1} a^-_{i_1 i_2}\right) \\ &= \lim_{n\to\infty} \sum_{\substack{i_2\in\mathbb{S}_2\\ i_2\leqslant n}}\left(\sum_{i_1\in\mathbb{S}_1} a_{i_1 i_2}\right) = \sum_{i_2\in\mathbb{S}_2}\left(\sum_{i_1\in\mathbb{S}_1} a_{i_1 i_2}\right).\end{aligned}$$

最后, 交换下标 1 和 2, 我们可得到相反求和次序的关系式.

6.2 概率建模

为了理解如何将上节的内容与概率理论相联系, 我们来考虑一下建模问题. 投掷一枚均匀的硬币 n 次, 取 $\Omega = \{0,1\}^n$, $\mathcal{F}$ 是 Ω 的所有子集的集合, 令 $\mu(\{\omega\}) = 2^{-n}$ 对任意的 $\omega \in \Omega$ 成立. 称它为 Kolmogorov 模型. 更一般地, 任何 Ω 上全测度为 1 的测度空间可以看做一个概率的模型, 我们把这个测度空间称为*概率空间*, 测度 μ 称为*概率测度*, 常用 P 来表示. 我们也将上节中的某些其他专业名词作相应的替换. 称 Ω 为*样本空间*, 它的元素称为*样本点*, $\mathcal{F}$ 的元素称为*事件*, 由一个事件指定的数字 P 称为该事件的*概率*. 另外, 一个可测映射被称为*随机变量*, 一般我们用 X 而不是 F 来表示. 当它取实值时, 它的积分就称为*期望*, 用 $\mathrm{E}[X]$ 来表示. 当需要确切表示时, 更准确的可以用 $\mathrm{E}^{\mathrm{P}}[X]$ 来表示 $\int X d\mathrm{P}$. 又$\mathrm{E}[X,A]$ 或 $\mathrm{E}^{\mathrm{P}}[X,A]$ 被用来表示 $\int_A X d\mathrm{P}$, 即 X 在事件 A 上的期望. 最后, 一个取值在可测空间 $(E,\mathcal{B})$ 里的*随机变量的分布*是 $(E,\mathcal{B})$ 上的概率测度 $X_*\mathrm{P}$, 定义为 $\mu(B) = \mathrm{P}(X \in B)$. 特别当 X 取实值时, 它的*分布函数* F_X 定义为 $F_X(x) = (X_*\mathrm{P})((-\infty, x]) = \mathrm{P}(X \leqslant x)$. 显然, F_X 是非降的. 由 (6.1.3) 可证 F_X 是右连续的, 即对任意的 $x \in \mathbb{R}$, $F_X(x) = \lim_{y\searrow x} F_X(y)$. 此外, $\lim_{x\searrow -\infty} F_X(x) = 0$, $\lim_{x\nearrow\infty} F_X(x) = 1$.

同时, 还有 $P(X < x) = F_X(x-) \equiv \lim\limits_{y \nearrow x} F_X(y)$, 由此可得, $P(X = x) = F_X(x) - F_X(x-)$ 即为 F_X 在 x 处的跳跃.

6.2.1. 无穷多次投掷均匀硬币的模型: 在测度空间的一般理论下, 只要样本空间是有限或可数的, 构建一个概率空间并不是分析的问题(而是组合的问题). 唯一的改变就是样本点的概率分布必须满足 $\sum\limits_{\omega\in\Omega} P(\{\omega\}) = 1$. 但是当 Ω 不可数时, 分析问题就出现了. 例如, 假设我们在投掷 n 次硬币后不是停止而是继续不停地投掷, 这就是一个无穷多次投掷硬币模型. 此时样本空间为 $\Omega = \{0,1\}^{\mathbb{Z}^+}$. 若 $A \subseteq \Omega$ 只依赖于前 n 次投掷硬币的结果, 那么 A 是一个可测事件且 A 的概率应该和在投掷 n 次硬币后停止的概率相同. 也就是说, 如果 $\Gamma \subseteq \{0,1\}^n$ 且[8] $A = \{\omega \in \Omega : (\omega(1), \cdots, \omega(n)) \in \Gamma\}$, 那么 $P(A)$ 应该等于 $2^{-n}\#\Gamma$.

继续无穷多次投掷硬币的例子, 可以看出 (参看 (6.1.3), 对于任一固定的 $\eta \in \Omega$, $P(\{\eta\})$ 等于

$$\lim_{n\to\infty} P(\{\omega : (\omega(1), \cdots, \omega(n)) = (\eta(1), \cdots, \eta(n))\}) = \lim_{n\to\infty} 2^{-n} = 0.$$

在这种情形, 当考虑每一个样本点的概率时, 由于它们的概率都为 0, 所以这样定义概率的方式没有反映任何东西. 实际上, 在 Ω 上是否存在一个具有所要求的性质的概率测度并不明显. 不过, 我们将证明这样的测度的确存在.

任取 $n \geqslant 1$, 令 $\Pi_n : \Omega \to \Omega_n \equiv \{0,1\}^n$ 表示由 $\omega \rightsquigarrow (\omega(1), \cdots, \omega(n))$ 给定的投影映射, 且令

$$\mathcal{A}_n = \{A \subseteq \Omega : A = \Pi_n^{-1}(\Pi_n(A))\}.$$

等价地, $A \in \mathcal{A}_n$ 当且仅当 A 依赖于前 n 个坐标, 其含义是: $\omega \in A$ 和 $\Pi_n(\omega') = \Pi_n(\omega) \implies \omega' \in A$. 其次, 在 $\mathcal{A}_n$ 上定义 P_n, 使得 $P_n(A) = 2^{-n}\#\Pi_n(A)$. 易知 $\mathcal{A}_n \subseteq \mathcal{A}_{n+1}$. 实际上, 若 $A \in \mathcal{A}_n$, 那么 $\Pi_{n+1}(A) = (\Pi_n(A) \times \{0,1\})$, 因此 $P_{n+1}(A) = P_n(A)$. 这样我们就能确切地在 $\mathcal{A} \equiv \bigcup\limits_{n=1}^{\infty} \mathcal{A}_n$ 上定义 P, 使得当 $A \in \mathcal{A}_n$ 时, $P(A) = P_n(A)$. 此外, 如果 $A, A' \in \mathcal{A}$ 互不相交, 那么通过选择适当的 n, 使得 $A, A' \in \mathcal{A}_n$, 就有 $P(A \bigcup A') = P_n(A \bigcup A') = P_n(A) + P_n(A') = P(A) + P(A')$.

---

[8] 为了便于理解, 可将 Ω 看做一个从 $\mathbb{Z}^+$ 到 $\{0,1\}$ 的映射组成的集合. 这样我们可以用 $\omega(n)$ 来表示 ω 的第 n 个坐标.

在进入下一步之前, 为了方便我们在 Ω 上引入一个距离, 使得 $\mathcal{A}$ 的元素都是开集. 也就是令

$$\rho(\omega,\omega') \equiv \sum_{n=1}^{\infty} 2^{-n}|\omega(n)-\omega'(n)|.$$

易证 ρ 是 Ω 上的一个距离. 实际上, 在这个距离下的收敛性相当于每一坐标的收敛性. 另外, 若 $\omega \in A \in \mathcal{A}_n$, 那么 $\rho(\omega,\omega') < 2^{-n} \implies \omega' \in A$. 所以, 每一个 $A \in \mathcal{A}$ 关于由 ρ 引出的拓扑是开集. 同时, 每一 $A \in \mathcal{A}$ 在 ρ-拓扑中都是闭的. 的确, 如果 $A \in \mathcal{A}_n$, $\{\omega_k\}_1^\infty \subseteq A$ 且 $\rho(\omega_k,\omega) \to 0$, 那么存在一个 $k \in \mathbb{Z}^+$ 使得 $\rho(\omega,\omega_k) < 2^{-n}$. 因此对于这个 k, $\Pi_n(\omega) = \Pi_n(\omega_k)$. 另一个我们需要的事实是, Ω 是 ρ-拓扑紧的. 为了证明它, 我们假设 $\{\omega_k\}_1^\infty$ 是 Ω 中的一个序列. 我们用对角线方法来找一收敛子列. 实际上, 因为对任意的 m, 或者对无穷多个 $k \in \mathbb{Z}^+, \omega_k(m) = 0$, 或者对无穷多个 $k \in \mathbb{Z}^+, \omega_k(m) = 1$. 所以可以利用归纳法去构造 $\{k_{m,\ell} : (m,\ell) \in (\mathbb{Z}^+)^2\}$ 使得 $\{k_{1,\ell} : \ell \in \mathbb{Z}^+\}$ 当有无限个 k, $\omega_k(1) = 0$ 时是 $\{k : \omega_k(1) = 0\}$ 的一个递增枚举, 而当只有有限个 k, $\omega_k(1) = 0$ 时是 $\{k : \omega_k(1) = 1\}$ 的一个递增枚举; 同样当 $m > 1$ 时, $\{k_{m,\ell} : \ell \in \mathbb{Z}^+\}$ 当有无限个 ℓ, $\omega_{k_{m-1,\ell}}(m) = 0$ 时是 $\{k_{m-1,\ell} : \omega_{k_{m-1,\ell}}(m) = 0\}$ 的一个递增枚举, 而当只有有限个 ℓ, $\omega_{k_{m-1,\ell}}(m) = 0$ 时是 $\{k_{m-1,\ell} : \omega_{k_{m-1,\ell}}(m) = 1\}$ 的一个递增枚举. 现在我们令 $k_\ell = k_{\ell,\ell}$, 那么可以证明 $\{\omega_{k_\ell} : \ell \in \mathbb{Z}^+\}$ 是 $\{\omega_k\}_1^\infty$ 的一个收敛子列. 最后我们还需要另外一个性质: 对任一开集 $G \subseteq \Omega$, 存在一个互不相交的集合序列 $\{A_m\}_1^\infty \subseteq \mathcal{A}$, 使得 $A_m \in \mathcal{A}_m, G = \bigcup\limits_1^\infty A_m$. 我们可以按下列步骤生成 $\{A_m\}_1^\infty$. 选取 A_1 是满足 $A \subseteq G$ 和 $A \in \mathcal{A}_1$ 的最大元素. 也就是说 A_1 是所有包含在 G 中的 $A \in \mathcal{A}_1$ 的并集. 然后, 如果 A_ℓ $(1 \leqslant \ell \leqslant m)$ 已经给定, 就令 A_{m+1} 是包含在 $G \setminus \bigcup\limits_1^m A_m$ 中的 $A \in \mathcal{A}_{m+1}$ 的最大者. 显然, 这些 A_m 都是互不相交的. 为证明它们覆盖 G, 假设 $\omega \in G$, 取 $n \geqslant 2$, 使得只要 $\rho(\omega',\omega) < 2^{-n+1}$, 就有 $\omega' \in G$. 这时 $A \equiv \{\omega' : \Pi_n(\omega') = \Pi_n(\omega)\} \subseteq G$, 并且, 要么 $\omega \in \bigcup\limits_1^{n-1} A_m$, 要么 $A \cap \bigcup\limits_1^{n-1} A_m = \varnothing$. 这时 $\omega \in A \subseteq A_n$.

有了上面的准备, 就可以继续我们的构造. 首先, 对于 $\Gamma \subseteq \Omega$, 定义 $\overline{\mathrm{P}}(\Gamma)$ 是 $\sum\limits_1^\infty \mathrm{P}(A_m)$ 在 Γ 的所有可数覆盖 $\{A_m\}_1^\infty \subseteq \mathcal{A}$ 上的下确界. 等价地, 因为 $\mathcal{A}$ 中的所有元素都是开的, 而一切开集都可以用 $\mathcal{A}$ 的可

数多个元的并来表示, 所以

$$\overline{\mathrm{P}}(\Gamma)=\inf\{\overline{\mathrm{P}}(G)\ :\ G\supseteq\Gamma,\ G\text{ 是开集}\}. \tag{6.2.1}$$

下面的引理包含了 $\overline{\mathrm{P}}$ 的某些基本性质.

6.2.2 引理. 对每个 $A\in\mathcal{A}$, $\overline{\mathrm{P}}(A)=\mathrm{P}(A)$. 更一般地, 对所有的 $\Gamma_1\subseteq\Gamma_2\subseteq\Omega$, $\overline{\mathrm{P}}(\Gamma_1)\leqslant\overline{\mathrm{P}}(\Gamma_2)$. 进一步, 若任意 $G\subseteq\Omega$ 是开的, $\{A_m\}_1^\infty$ 是由 $\mathcal{A}$ 中的互不相交的元素组成的 G 的一个可数的精确覆盖, 则 $\overline{\mathrm{P}}(G)=\sum\limits_1^\infty\mathrm{P}(A_m)$. 并且, 一般地, 对于 Ω 的任意子集序列 $\{\Gamma_k\}_1^\infty$, 有

$$\overline{\mathrm{P}}\Big(\bigcup_1^\infty\Gamma_k\Big)\leqslant\sum_1^\infty\overline{\mathrm{P}}\Big(\Gamma_k\Big). \tag{6.2.3}$$

最后, 如果 F 和 F' 是 Ω 的两个互不相交的闭子集, 则 $\overline{\mathrm{P}}(F\bigcup F')=\overline{\mathrm{P}}(F)+\overline{\mathrm{P}}(F')$.

证明: 因为任何 Γ_2 的覆盖也是 Γ_1 的覆盖, 第二个结论是显然的. 下面假设 $A\in\mathcal{A}$. 显然, $\overline{\mathrm{P}}(A)\leqslant\mathrm{P}(A)$. 另一方面, 如果 $\{A_m\}_1^\infty$ 是由 $\mathcal{A}$ 中的元素组成的 A 的一个覆盖, 则由紧集的 Heine-Borel 性质, 我们可以找出 (记住 A 是闭的, 因此是紧的) 一个 n 使得 $A\subseteq\bigcup\limits_1^n A_m$. 现在选取 N 使得 $\{A\}\bigcup\{A_m\}_1^n\subseteq\mathcal{A}_N$. 于是

$$\mathrm{P}(A)=\mathrm{P}_N(A)\leqslant\sum_1^n\mathrm{P}_N(A_m)=\sum_1^n\mathrm{P}(A_m)\leqslant\sum_1^\infty\mathrm{P}(A_m).$$

因此我们可以得出 $\mathrm{P}(A)\leqslant\overline{\mathrm{P}}(A)$.

下面令 G 是开的, 假设 $\{A_m\}_1^\infty$ 是由 $\mathcal{A}$ 中的互不相交的元素组成的 G 的一个精确覆盖. 由定义, $\overline{\mathrm{P}}(G)\leqslant\sum\limits_1^\infty\mathrm{P}(A_m)$. 另一方面, 因为 $G\supseteq\bigcup\limits_1^n A_m\in\mathcal{A}$, 我们知道

$$\begin{aligned}\overline{\mathrm{P}}(G)\geqslant\overline{\mathrm{P}}\Big(\bigcup_1^n A_m\Big)&=\mathrm{P}\Big(\bigcup_1^n A_m\Big)\\&=\mathrm{P}_N\Big(\bigcup_1^n A_m\Big)=\sum_1^n\mathrm{P}_N(A_m)=\sum_1^n\mathrm{P}(A_m),\end{aligned}$$

这里选取 N 使得 $\{A_m\}_1^n\subseteq\mathcal{A}_N$. 因此, 令 $n\to\infty$, 我们得到想要的结论.

现在假设 $\{\Gamma_k\}_1^\infty$ 是 Ω 的一列子集. 给定 $\epsilon>0$, 对于每一个 $k\in\mathbb{Z}^+$ 选取一个 Γ_k 的可数覆盖 $\{A_{k,\ell}:\ell\in\mathbb{Z}^+\}\subseteq\mathcal{A}$, 使得 $\sum\limits_{\ell}\mathrm{P}(A_{k,\ell})\leqslant\overline{\mathrm{P}}(\Gamma_k)+2^{-k}\epsilon$. 于是 $\{A_{k,\ell}:(k,\ell)\in(\mathbb{Z}^+)^2\}\subseteq\mathcal{A}$ 是 $\bigcup\limits_k\Gamma_k$ 的可数覆盖. 因此 (由可数测度空间的 Fubini 定理)

$$\begin{aligned}\overline{\mathrm{P}}\Big(\bigcup_k\Gamma_k\Big)&\leqslant\sum_{(k,\ell)\in(\mathbb{Z}^+)^2}\mathrm{P}(A_{k,\ell})=\sum_{k\in\mathbb{Z}^+}\Big(\sum_{\ell\in\mathbb{Z}^+}\mathrm{P}(A_{k,\ell})\Big)\\&\leqslant\sum_{k\in\mathbb{Z}^+}\overline{\mathrm{P}}(\Gamma_k)+\epsilon.\end{aligned}$$

得证 $\overline{\mathrm{P}}(\bigcup\limits_k\Gamma_k)\leqslant\sum\limits_k\overline{\mathrm{P}}(\Gamma_k)$.

最后, 给定 Ω 的两个互不相交的闭子集 F 和 F', 上面的证明推出 $\overline{\mathrm{P}}(F\bigcup F')\leqslant\overline{\mathrm{P}}(F)+\overline{\mathrm{P}}(F')$. 为了得到相反的不等式, 首先注意到, 因为它们是紧的, 所以存在一个 $N\geqslant 2$ 使得 $\rho(\omega,\omega')>2^{-N+1}$ 对于所有的 $\omega\in F$ 和 $\omega'\in F'$ 成立. 因此, 如果 $B\equiv\Pi_N^{-1}(\Pi_N(F))$ 和 $B'\equiv\Pi_N^{-1}(\Pi_N(F'))$, 则 $F\subseteq B\in\mathcal{A}$, $F'\subseteq B'\in\mathcal{A}$, 并且 $B\bigcap B'=\varnothing$, 后者是因为由 $\eta\in B\bigcap B'$ 可推出存在 $\omega\in F$ 和 $\omega'\in F'$ 使得 $\rho(\omega,\omega')\leqslant\rho(\omega,\eta)+\rho(\eta,\omega')\leqslant 2^{-N+1}$. 现在假设 $\{A_m\}_1^\infty\subseteq\mathcal{A}$ 是 $F\bigcup F'$ 的一个可数覆盖, 记 $B_m=B\bigcap A_m$ 和 $B'_m=B'\bigcap A'_m$. 则 $\{B_m\}_1^\infty\subseteq\mathcal{A}$ 和 $\{B'_m\}_1^\infty\subseteq\mathcal{A}$ 分别是 F 和 F' 的可数覆盖. 另外, 由于 B_m 和 B'_m 是 $\mathcal{A}$ 中两个互不相交的集合, 所以 $\mathrm{P}(A_m)\geqslant\mathrm{P}(B_m\bigcup B'_m)=\mathrm{P}(B_m)+\mathrm{P}(B'_m)$. 因此

$$\sum_m\mathrm{P}(A_m)\geqslant\sum_m\mathrm{P}(B_m)+\sum_m\mathrm{P}(B'_m)\geqslant\overline{\mathrm{P}}(F)+\overline{\mathrm{P}}(F').\ \square$$

在我们的构造过程中最后一点要说明的是 Ω 中什么样的子集 Γ 是可测的. 尽管理由并不非常清楚, 但是如果对于每个 $\epsilon>0$, 存在一个开集 $G\supseteq\Gamma$ 使得 $\overline{\mathrm{P}}(G\backslash\Gamma)\leqslant\epsilon$, 我们说 $\Gamma\subseteq\Omega$ 是可测的, 并且记为 $\Gamma\in\bar{\mathcal{B}}$. 显然, 每个开集是可测的. 同时, 容易知道如果 $\overline{\mathrm{P}}(\Gamma)=0$, 则 $\Gamma\in\bar{\mathcal{B}}$. 实际上, 由 (6.2.1) 式, 假如 $\overline{\mathrm{P}}(\Gamma)=0$, 则对于每个 $\epsilon>0$, 我们可以找到一个开集 $G\supseteq\Gamma$, 使得 $\overline{\mathrm{P}}(G\backslash\Gamma)\leqslant\overline{\mathrm{P}}(G)\leqslant\epsilon$. 然而, 每个闭集 $F\subseteq\Omega$ 是可测的这一事实并不明显. 为了明白这一点, 令 $\epsilon>0$, 选取一个开集 $G\supseteq F$ 使得 $\overline{\mathrm{P}}(G)\leqslant\overline{\mathrm{P}}(F)+\epsilon$. 则 $G\backslash F$ 是开的, 所以我们能写 $G\backslash F=\bigcup\limits_m A_m$, 这里 A_m 是 $\mathcal{A}$ 中互不相交的元. 因此, 由引理 6.2.2,

$\overline{\mathrm{P}}(G\backslash F) = \sum_1^\infty \mathrm{P}(A_m)$. 另一方面, 对于任何 $n \geqslant 1$, $B_n \equiv \bigcup_1^n A_m$ 是 $G\backslash F$ 的一个闭子集, 这样, 由引理 6.2.2 的第一部分和最后部分,

$$\sum_1^n \mathrm{P}(A_m) = \mathrm{P}(B_n) = \overline{\mathrm{P}}(B_n) = \overline{\mathrm{P}}(F \bigcup B_n) - \overline{\mathrm{P}}(F) \leqslant \overline{\mathrm{P}}(G) - \overline{\mathrm{P}}(F) \leqslant \epsilon.$$

因此, $\overline{\mathrm{P}}(G\backslash F) = \sum_1^\infty \mathrm{P}(A_m) \leqslant \epsilon$.

至此我们知道开集、闭集以及 $\overline{\mathrm{P}}$ 测度为零的集合都是可测的, 我们能够证明 $\bar{\mathcal{B}}$ 是一个 σ-代数. 为此, 首先假设 $\{\Gamma_k\}_1^\infty \subseteq \bar{\mathcal{B}}$, 给定 $\epsilon > 0$, 选取开集 $G_k \supseteq \Gamma_k$ 使得 $\overline{\mathrm{P}}(G_k\backslash\Gamma_k) \leqslant 2^{-k}\epsilon$. 那么 $G \equiv \bigcup_1^\infty G_k$ 是开的, $G \supseteq \Gamma \equiv \bigcup_1^\infty \Gamma_k$, 且

$$\overline{\mathrm{P}}(G\backslash F) \leqslant \overline{\mathrm{P}}\Big(\bigcup_1^\infty (G_k\backslash\Gamma_k)\Big) \leqslant \sum_1^\infty \mathrm{P}(G_k\backslash\Gamma_k) \leqslant \epsilon.$$

因此, $\Gamma \in \bar{\mathcal{B}}$, 所以 $\bar{\mathcal{B}}$ 对可数并运算封闭. 为了证明它对余集运算也封闭, 给定 $\Gamma \in \bar{\mathcal{B}}$, 对于每个 $n \geqslant 1$, 选取一个开集 $G_n \supseteq \Gamma$ 使得 $\overline{\mathrm{P}}(G_n\backslash\Gamma) \leqslant \frac{1}{n}$. 则 $D \equiv \bigcap_1^\infty G_n \supseteq \Gamma$, 且对于所有的 $n \geqslant 1$, $\overline{\mathrm{P}}(D\backslash\Gamma) \leqslant \overline{\mathrm{P}}(G_n\backslash\Gamma) \leqslant \frac{1}{n}$. 因此 $\overline{\mathrm{P}}(D\backslash\Gamma) = 0$. 所以 $D\backslash\Gamma \in \bar{\mathcal{B}}$. 现在记 $F_n = G_n^c$, $C = \bigcup_1^\infty F_n$. 因为每个 F_n 是闭的, 从而是可测的, 所以 $C \in \bar{\mathcal{B}}$. 注意到 $\Gamma^c = C \bigcup (D\backslash\Gamma) \in \bar{\mathcal{B}}$, 证明完成. □

6.2.4 定理. *沿用上面的记号, $\mathcal{B}_\Omega = \sigma(\mathcal{A})$ 和 $\Gamma \in \bar{\mathcal{B}}$ 当且仅当存在 $C, D \in \mathcal{B}_\Omega$ 使得 $C \subseteq \Gamma \subseteq D$ 及 $\overline{\mathrm{P}}(D\backslash C) = 0$. 此外仍用 P 表示 $\overline{\mathrm{P}}$ 在 $\bar{\mathcal{B}}$ 的限制 $\overline{\mathrm{P}} \upharpoonright \bar{\mathcal{B}}$. 则 $(\Omega, \bar{\mathcal{B}}, \mathrm{P})$ 是一个概率空间, 满足 $\mathrm{P} \upharpoonright \mathcal{A}_n = \mathrm{P}_n$, $n \geqslant 1$.*

证明: 我们已经证明了 $\bar{\mathcal{B}}$ 是一个 σ-代数. 因为 $\mathcal{A}$ 中的所有元素都是开的, 而每一个开集可以写作可数个 $\mathcal{A}$ 中元素的并, 所以 $\mathcal{B}_\Omega = \sigma(\mathcal{A})$. 为了根据 $\mathcal{B}_\Omega$ 证明 $\bar{\mathcal{B}}$ 的性质, 首先指出, 因为开集一定在 $\bar{\mathcal{B}}$ 中, 所以 $\mathcal{B}_\Omega \subseteq \bar{\mathcal{B}}$. 此外, 因为 $\overline{\mathrm{P}}$-测度为零的集合也在 $\bar{\mathcal{B}}$ 中, 因此对于满足 $C, D \in \mathcal{B}_\Omega$, $\overline{\mathrm{P}}(D\backslash C) = 0$ 的 $C \subseteq \Gamma \subseteq D$, 由 $\overline{\mathrm{P}}(\Gamma\backslash C) \leqslant \overline{\mathrm{P}}(D\backslash C) = 0$, 得到 $\Gamma = C \bigcap (\Gamma\backslash C) \in \bar{\mathcal{B}}$. 反过来, 如果 $\Gamma \in \bar{\mathcal{B}}$, 我们可以找到开集序列 $\{G_n\}_1^\infty$ 和 $\{H_n\}_1^\infty$, 使得 $G_n \supseteq \Gamma$, $H_n \supseteq \Gamma^c$, 且 $\overline{\mathrm{P}}(G_n\backslash\Gamma) \vee \overline{\mathrm{P}}(H_n\backslash\Gamma^c) \leqslant \frac{1}{n}$. 因此, 如果 $D = \bigcap_1^\infty G_n$ 且 $C = \bigcup_1^\infty H_n^c$, 则 $C, D \in \mathcal{B}_\Omega$, $C \subseteq \Gamma \subseteq D$, 且

对于所有的 $n \geqslant 1$,

$$\overline{\mathrm{P}}(D \backslash C) \leqslant \overline{\mathrm{P}}(G_n \backslash H_n^c) \leqslant \overline{\mathrm{P}}(G_n \backslash \Gamma) + \overline{\mathrm{P}}(H_n \backslash \Gamma^c) \leqslant \frac{2}{n}.$$

因此 $\overline{\mathrm{P}}(D \backslash C) = 0$.

余下只要验证 $\overline{\mathrm{P}}$ 在 $\bar{\mathcal{B}}$ 上是可数可加的. 为此令 $\{\Gamma_k\}_1^\infty \subseteq \bar{\mathcal{B}}$ 是一列互不相交的集合, 记 $\Gamma = \bigcup_1^\infty \Gamma_k$. 由引理 6.2.2 的第二部分, 我们知道 $\overline{\mathrm{P}}(\Gamma) \leqslant \sum_1^\infty \overline{\mathrm{P}}(\Gamma_k)$. 为了得到相反的不等式, 给定 $\epsilon > 0$, 对于每个 k, 选取开集 $G_k \supseteq \Gamma_k^c$ 使得 $\overline{\mathrm{P}}(G_k \backslash \Gamma_k^c) \leqslant 2^{-k}\epsilon$. 则每个 $F_k = G_k^c$ 是闭的, 且

$$\overline{\mathrm{P}}(\Gamma_k) \leqslant \overline{\mathrm{P}}(F_k) + \overline{\mathrm{P}}(G_k \backslash \Gamma_k^c) \leqslant \mathrm{P}(F_k) + 2^{-k}\epsilon.$$

另外, 因为 Γ_k 是互不相交的, F_k 也是互不相交的, 所以, 由引理 6.2.2 的最后部分, 我们知道, 对于所有的 $n \geqslant 1$ 有

$$\overline{\mathrm{P}}(\Gamma) \geqslant \overline{\mathrm{P}}\Big(\bigcup_1^n F_k\Big) = \sum_1^n \overline{\mathrm{P}}(F_k),$$

并且

$$\sum_1^\infty \overline{\mathrm{P}}(\Gamma_k) \leqslant \sum_1^\infty \overline{\mathrm{P}}(F_k) + \epsilon \leqslant \overline{\mathrm{P}}(\Gamma) + \epsilon. \quad \square$$

6.3 独立随机变量

在 Kolmogorov 模型下, 描述独立性的最好方法是通过 σ-代数. 如果 $(\Omega, \mathcal{F}, \mathrm{P})$ 是一个概率空间, $\mathcal{F}_1$ 和 $\mathcal{F}_2$ 是 $\mathcal{F}$ 的子 σ-代数 (也即, 是 $\mathcal{F}$ 的子集, 且为 σ-代数). 如果对于所有的 $\Gamma_1 \in \mathcal{F}_1$ 和 $\Gamma_2 \in \mathcal{F}_2$, $\mathrm{P}(\Gamma_1 \bigcap \Gamma_2) = \mathrm{P}(\Gamma_1)\mathrm{P}(\Gamma_2)$, 我们说 $\mathcal{F}_1$ 和 $\mathcal{F}_2$ 是**独立的**. 考虑 $A_1, A_2 \in \mathcal{F}$, 对于 $i \in \{1,2\}$, 定义 $\mathcal{F}_i = \sigma(\{A_i\}) = \{\varnothing, A_i, A_i^c, \Omega\}$. 那么容易验证, 当 (用初等概率论的术语) “A_1 独立于 A_2”: $\mathrm{P}(A_1 \bigcap A_2) = \mathrm{P}(A_1)\mathrm{P}(A_2)$ 时, $\mathcal{F}_1$ 独立于 $\mathcal{F}_2$.

独立性的概念可用于随机变量上. 令 $\{X_\alpha : \alpha \in \mathcal{I}\}$ 是 $(\Omega, \mathcal{F}, \mathrm{P})$ 上的随机变量族, 如果对 $\mathcal{I}$ 中的任意两个互不相交的子集 $\mathcal{J}_1$ 和 $\mathcal{J}_2$, σ-代数 $\sigma(\{X_\alpha : \alpha \in \mathcal{J}_1\})$ 和 $\sigma(\{X_\alpha : \alpha \in \mathcal{J}_2\})$ 相互独立, 那么称 $\{X_\alpha : \alpha \in \mathcal{I}\}$ 中随机变量是相互独立的. 我们可以利用定理 6.1.6 来证明这一定义等价于, 如果 X_α 取值于可测空间 $(\mathrm{E}_\alpha, \mathcal{B}_\alpha)$, 那么对由

$\mathcal{I}$ 中的不同元素组成的有限子集 $\{\alpha_m\}_1^n$ 和任意选取的 $B_{\alpha_m} \in \mathcal{B}_{\alpha_m}$, $1 \leqslant m \leqslant n$, 有

$$\mathrm{P}(X_{\alpha_m} \in B_{\alpha_m}, 1 \leqslant m \leqslant n) = \prod_1^n \mathrm{P}(X_{\alpha_m} \in B_{\alpha_m}).$$

这个定义的一个优点是, 如果 $\{X_\alpha : \alpha \in \mathcal{I}\}$ 是相互独立的且对于每个 $\alpha \in \mathcal{I}$, F_α 是一个 X_α 值域上的可测映射, 则 $\{F_\alpha(X_\alpha) : \alpha \in \mathcal{I}\}$ 也是相互独立的. 最后, 由简单函数开始, 我们可以证明如果 $\{X_m\}_1^n$ 是相互独立的, 且对 $1 \leqslant m \leqslant n$, f_m 是定义于 X_m 值域上的 $\mathbb{R}$ 值可测函数, 那么只要 f_m 都有界或者都非负, 就有

$$\mathrm{E}[f_1(X_1) \cdots f_n(X_n)] = \prod_1^n \mathrm{E}[f_m(X_m)].$$

6.3.1. 独立随机变量族的存在性: 在上一节, 我们构造了一族可数个相互独立的随机变量. 也即, 如果 $(\Omega, \mathcal{B}, \mathrm{P})$ 是定理 6.2.4 中讨论的概率空间, $X_m(\omega) = \omega(m)$ 是 ω 的第 m 个坐标, 则对于任何 $n \geqslant 1$ 和 $(\eta_1, \cdots, \eta_n) \in \{0,1\}$, $\mathrm{P}(X_m = \eta_m, 1 \leqslant m \leqslant n) = 2^{-n} = \prod\limits_1^n \mathrm{P}(X_m = \eta_m)$. 因此, 随机变量 $\{X_m\}_1^\infty$ 是相互独立的. 只取两个值的相互独立的随机变量, 叫做 *Bernoulli* 随机变量.

正如我们即将看到的, Bernoulli 随机变量可以用来构造许多其他相互独立的随机变量族. 这种构造的关键被包含在下面的引理中.

6.3.1 引理. 任给一族相互独立的 $\{0,1\}$ 值的 *Bernoulli* 随机变量 $\{B_m\}_1^\infty$, 满足 $\mathrm{P}(B_m = 0) = \frac{1}{2} = \mathrm{P}(B_m = 1)$, $m \in \mathbb{Z}^+$. 记 $U = \sum\limits_1^\infty 2^{-m} B_m$. 则 U 是 $[0,1)$ 上的均匀分布变量. 也就是说, 如果 $u < 0$, $\mathrm{P}(U \leqslant u)$ 是 0; 如果 $u \in [0,1)$, $\mathrm{P}(U \leqslant u)$ 是 u; 如果 $u \geqslant 1$, $\mathrm{P}(U \leqslant u)$是 1.

证明: 给定 $N \geqslant 1$ 和 $0 \leqslant n < 2^N$, 我们要证明

$$\mathrm{P}(n2^{-N} < U \leqslant (n+1)2^{-N}) = 2^{-N}. \tag{$*$}$$

为此, 注意到 $n2^{-N} < U \leqslant (n+1)2^{-N}$ 当且仅当 $\sum\limits_1^N 2^{-m} B_m = (n+1)2^{-N}$ 且对 $m > N$, $B_m = 0$, 或者 $\sum\limits_1^N 2^{-m} B_m = n2^{-N}$ 且对某 $m > N$, $B_m = 1$. 因此, 由于 $\mathrm{P}(B_m = 0,$ 对所有 $m > N) = 0$, $(*)$ 的左边等于

事件 $\sum\limits_1^N 2^{-m}B_m = n2^{-N}$ 的概率. 然而, 一些初等的考虑表明, 对任意的 $0 \leqslant n < 2^N$, 存在唯一的 $(\eta_1, \cdots, \eta_N) \in \{0,1\}^N$, 满足 $\sum\limits_1^N 2^{-m}\eta_m = n2^{-N}$. 因此有

$$\mathrm{P}\left(\sum_1^N 2^{-m}B_m = n2^{-N}\right) = \mathrm{P}(B_m = \eta_m, 1 \leqslant m \leqslant N) = 2^{-N}.$$

这证明 $(*)$ 式成立, 剩下的就简单了. 因为 $\mathrm{P}(U=0) = \mathrm{P}(B_m = 0, m \geqslant 1) = 0$, 由 $(*)$ 式可知, 对任意 $1 \leqslant k \leqslant 2^N$,

$$\mathrm{P}(U \leqslant k2^{-N}) = \sum_{m=0}^{k-1} \mathrm{P}(m2^{-N} < U \leqslant (m+1)2^{-N}) = k2^{-N}.$$

由此, 并注意到 $u \rightsquigarrow \mathrm{P}(U \leqslant u)$ 是右连续的, 易知, 对任意的 $u \in [0,1)$, $F_U(u) = u$. 最后, 因为 $\mathrm{P}(U \in [0,1]) = 1$, 这就完成了定理的证明. □

现在, 令 $\mathcal{I}$ 为非空的有限或可数无穷集. 那么 $\mathcal{I} \times \mathbb{Z}^+$ 是可数的, 所以我们可以构造从 $\mathcal{I} \times \mathbb{Z}^+$ 到 $\mathbb{Z}^+$ 的 $1-1$ 满射: $(\alpha, n) \rightsquigarrow N(\alpha, n)$. 然后, 对每个 $\alpha \in \mathcal{I}$, 定义 $\omega \in \Omega = \{0,1\}^{\mathbb{Z}^+} \mapsto \boldsymbol{X}_\alpha(\omega) \in \Omega$, 使得 $\boldsymbol{X}_\alpha(\omega)$ 的第 n 个分量是 $\omega(N(\alpha,n))$. 那么作为定理 6.2.4 中概率空间 $(\Omega, \bar{\mathcal{B}}, \mathrm{P})$ 上的随机变量, $\{\boldsymbol{X}_\alpha : \alpha \in \mathcal{I}\}$ 是相互独立的, 而且每个都有分布 P. 因此, 如果 $\Phi : \Omega \to [0,1]$ 是如下的连续映射

$$\Phi(\eta) \equiv \sum_{m=1}^{\infty} 2^{-m}\eta(m), \ \eta \in \Omega,$$

并且如果 $U_\alpha \equiv \Phi(\boldsymbol{X}_\alpha)$, 那么随机变量 $\{U_\alpha \equiv \Phi(\boldsymbol{X}_\alpha) : \alpha \in \mathcal{I}\}$ 是相互独立的, 且由引理 6.3.1, 每个都服从 $[0,1]$ 上均匀分布.

这一节的最后一步是将前面的构造和熟知的事实 (任何 $\mathbb{R}$ 值随机变量都可以通过均匀随机变量来表示) 结合起来. 更确切地说, 考虑一个映射 $F : \mathbb{R} \to [0,1]$, 如果 F 是非减、右连续的, 并且在 $-\infty$ 趋于 0, 在 ∞ 趋于 1, 则称为**分布函数**. 给定这样一个 F, 定义

$$F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geqslant u\}, \ u \in [0,1].$$

注意到, 由右连续性, $F(x) \geqslant u \iff F^{-1}(u) \leqslant x$. 因此, 如果 U 服从 $[0,1]$ 上的均匀分布, 那么 $F^{-1}(U)$ 就是分布函数为 F 的随机变量.

6.3.2 定理. 令 $(\Omega, \bar{\mathcal{B}}, \mathrm{P})$ 是定理 6.2.4 中的概率空间. 任给有限或可数无限的指标集 $\mathcal{I}$ 和分布函数集 $\{F_\alpha : \alpha \in \mathcal{I}\}$, 则存在 $(\Omega, \bar{\mathcal{B}}, \mathrm{P})$ 上的相互独立的随机变量 $\{X_\alpha : \alpha \in \mathcal{I}\}$ 使得对每个 $\alpha \in \mathcal{I}$, F_α 是 X_α 的分布函数.

6.4 条件概率和条件期望

正如对独立性的讨论那样, σ-代数是 Kolmogorov 定义 "条件" 的关键. 给定概率空间 $(\Omega, \mathcal{F}, \mathrm{P})$, 一个子 σ-代数 Σ 和一个非负或可积的随机变量 X, Kolmogorov 称随机变量 X_Σ 是 X 给定 Σ 下的*条件期望*, 如果 X_Σ 是非负或可积的、关于 Σ 可测 (即 $\sigma(\{X_\Sigma\}) \subseteq \Sigma$) 的随机变量, 并且对所有的 $\Gamma \in \Sigma$,

$$\mathrm{E}[X_\Sigma, \Gamma] = \mathrm{E}[X, \Gamma]. \tag{6.4.1}$$

当 X 是集合 $B \in \mathcal{F}$ 的示性函数时, 条件期望就改称为*条件概率*.

为了理解该定义是初等概率教程中条件期望定义的推广, 我们从 Σ 是平凡 σ-代数 $\{\varnothing, \Omega\}$ 开始考虑. 因为只有常值随机变量关于 $\{\varnothing, \Omega\}$ 是可测的, 显然, X 的唯一条件期望是 $\mathrm{E}[X]$. 下面假设 $\Sigma = \sigma(\{A\}) = \{\varnothing, A, A^c, \Omega\}$, 其中 $A \in \mathcal{F}$, 且$\mathrm{P}(A) \in (0,1)$. 在这种情形下, 容易验证, 对任意的 $B \in \mathcal{F}$,

$$\omega \rightsquigarrow \frac{\mathrm{P}(B \bigcap A)}{\mathrm{P}(A)} \mathbf{1}_A(\omega) + \frac{\mathrm{P}(B \bigcap A^c)}{\mathrm{P}(A^c)} \mathbf{1}_{A^c}(\omega)$$

是 B 给定 Σ 下的条件概率. 也就是, 在初等概率论中所称的 "给定事件 A, B 的条件概率" $\mathrm{P}(B|A) = \dfrac{\mathrm{P}(B \bigcap A)}{\mathrm{P}(A)}$, 在这里就是映射 $\omega \rightsquigarrow \mathrm{P}(B|\Sigma)(\omega)$ 在 A 上的值. 更一般地, 如果 Σ 是由 Ω 的一个有限或可数划分 $\mathcal{P} \subseteq \mathcal{F}$ 生成, 则对任意非负或可积的随机变量 X,

$$\omega \rightsquigarrow \sum_{\{A \in \mathcal{P} : \mathrm{P}(A) > 0\}} \frac{\mathrm{E}[X, A]}{\mathrm{P}(A)} \mathbf{1}_A(\omega)$$

是在给定 Σ 下 X 的条件期望.

当然, Kolmogorov 的定义提出了两个实质性的问题: 存在性和唯一性. 一般情形的存在性可以有很多种方法证明. 比如, 当 $\mathrm{E}[X^2] < \infty$ 时, 容易看出, 正如 $\mathrm{E}[X]\mathbf{1}$ 是 $\mathrm{E}[(X - X')^2]$ 在所有常值随机变量 X' 中

的最小值一样, X_Σ 也必定是 $\mathrm{E}[(X-X')^2]$ 在所有 Σ-可测的随机变量 X' 中的最小值. 用这种方法, 存在性问题和在所有均方可积随机变量组成的空间中的正交投影问题能够联系起来, 虽然这些内容超出了本书的范围, 但有关这类投影的结果对那些学过 Hilbert 空间的人来说应该是熟悉的.

另一方面, 唯一性比起存在性更简单, 也更微妙. 事实上, 这里没有通常单纯的唯一性结果, 因为一般情况下, 我们有无数种方式选择 X_Σ.[9] 而每个选择与其他任何选择最多只在一个零测集上不同. 为了说明这一点, 假设 X'_Σ 是满足 (6.4.1) 式的另一个非负或可积的随机变量, 则 $A=\{X'_\Sigma > X_\Sigma\}\in\Sigma$, 而且仅当 $\mathrm{P}(A)=0$ 时, (6.4.1) 才成立. 类似地, $\mathrm{P}(X_\Sigma > X'_\Sigma)=0$, 所以 $\mathrm{P}(X_\Sigma \neq X'_\Sigma)=0$.

尽管刚才讨论的唯一性问题有些含糊, 但通常我们尽可能地忽略这一点, 并且好像随机变量关于给定的 σ-代数只有一个条件期望似的. 就此而言, X 关于给定 Σ 下的条件期望的标准记号是 $\mathrm{E}[X|\Sigma]$, 或者当 $X=\mathbf{1}_B$ 时, 条件期望记为 $\mathrm{P}(B|\Sigma)$, 这正是我们在本书前几章中采用的记号.

6.4.1. 关于随机变量的条件运算: 在本书中, 基本上所有的条件运算都是关于 $\Sigma=\sigma(\mathfrak{F})$ 的 (参看 §6.1.4), 其中 $\mathfrak{F}$ 是 $(\Omega,\mathcal{F},\mathrm{P})$ 上的一族可测函数. 当 Σ 具有这种形式时, 随机变量 X 的条件期望是一个 $\mathfrak{F}$ 中的函数的可测函数. 例如, 如果 $\mathfrak{F}=\{F_1,\cdots,F_n\}$, 其中 F_m 都取值于一个可数空间 $\mathbb{S}$, 则 X 关于给定的 $\sigma(\mathfrak{F})$ 的条件期望 $\mathrm{E}[X|\sigma(\mathfrak{F})]$ 具有形式 $\Phi(F_1,\cdots,F_n)$, 其中根据 $\mathrm{P}(F_1=i_1,\cdots,F_n=i_n)$ 是正的或 0, $\Phi(i_1,\cdots,i_n)$ 等于

$$\frac{\mathrm{E}[X,F_1=i_1,\cdots,F_n=i_n]}{\mathrm{P}(F_1=i_1,\cdots,F_n=i_n)} \quad \text{或} \quad 0.$$

为了强调关于 $\sigma(\mathfrak{F})$ 的条件运算得到的是 $\mathfrak{F}$ 的一个函数, 我们用记号 $\mathrm{E}[X|\mathfrak{F}]$ 或 $\mathrm{P}(B|\mathfrak{F})$ 代替 $\mathrm{E}[X|\sigma(\mathfrak{F})]$ 或 $\mathrm{P}(B|\sigma(\mathfrak{F}))$.

为了给出上述讨论的更具体的例子, 首先假设 X 和 Y 是取值于某个可数空间 $\mathbb{S}$ 的独立随机变量, 令 $Z=F(X,Y)$, 其中 $F:\mathbb{S}^2\to\mathbb{R}$ 是有界的. 那么

$$\mathrm{E}[Z|X]=v(X), \quad \text{其中 } v(i)=\mathrm{E}[F(i,Y)], \quad i\in\mathbb{S}. \tag{6.4.2}$$

[9] 这种非唯一性就是在英文词 “conditional expectation” 前用冠词 “a”, 而不用 “the” 的原因.

Markov 链的讨论提供了一个不平凡的例子. 在第二章, 我们曾用下面的方程表示 Markov 性,

$$\mathrm{P}(X_{n+1}=j|X_0,\cdots,X_n)=(\boldsymbol{P})_{X_nj},$$

它将条件运算写成了赋予条件的随机变量的函数, 即 $(i_0,\cdots,i_n)\rightsquigarrow(\boldsymbol{P})_{i_nj}$. (当然, Markov 性的显著特征是, 这个函数只依赖于 i_n, 而不是 $(i_0,\cdots,i_{n-1})$.) 类似地, 当讨论连续时间参数的 Markov 过程时, 我们记

$$\mathrm{P}(X(t)=j|X(\sigma),\sigma\in[0,s])=(\boldsymbol{P}(t-s))_{X(s)j},$$

这再次说明, 经条件运算所得的结果是赋予条件的随机变量的函数.

符　　号

记号	含义	见
$\mathbb{Z}$ 和 $\mathbb{Z}^+$	所有整数的集合和所有正整数的集合	
$\mathbb{N}$	非负整数的集合	
$\#S$	集合 S 中元素的个数	
A^c	集合 A 的余集 (补集)	
$\mathbf{1}_A$	集合 A 的示性函数: 当 x 属于 A 时, $\mathbf{1}_A(x) = 1$, 当 x 不属于 A 时, $\mathbf{1}_A(x) = 0$	
$F \upharpoonright S$	函数 F 限制于集 S 上	
$a \wedge b$ 和 $a \vee b$	a 与 b 的最小值和最大值, $a, b \in \mathbb{R}$	
a^+ 和 a^-	$a \in \mathbb{R}$ 的正部 $a \vee 0$ 和负部 $(-a) \vee 0$	
$i \to j$ 和 $i \leftrightarrow j$	状态 j 可由 i 到达和状态 j 与 i 连通	§3.1
δ_{ij}	Kronecker delta 函数: δ_{ij} 取 1 或 0, 取决于 i 是否等于 j	
$\boldsymbol{\delta}_i$	$\mathbb{S}$ 中点 i 的单点测度: $(\boldsymbol{\delta}_i)_j = \delta_{ij}$, $j \in \mathbb{S}$	
$\mathrm{E}[X, A]$	X 在事件 A 上的期望值	§6.2
$\mathrm{E}[X\|A]$ 和 $\mathrm{E}[X\|\Sigma]$	在给定事件 A 和 σ-代数 Σ 下, X 的条件期望值	§6.4
$\langle\varphi\rangle_{\boldsymbol{\pi}}$	$\sum_{i\in\mathbb{S}} \varphi(i)(\boldsymbol{\pi})_i$ 的另一个记号	(5.1.4)
$\langle\varphi, \psi\rangle_{\boldsymbol{\pi}}$ 和 $\\|\varphi\\|_{2,\boldsymbol{\pi}}$	在 $L^2(\boldsymbol{\pi})$ 空间中两者的内积和范数	§5.1.2

续表

记号	含义	见
$\mathrm{Stat}(\boldsymbol{P})$	转移概率矩阵 $\boldsymbol{P}$ 的平稳分布的集合	§3.2.3
$\\|\boldsymbol{\mu}\\|_v$	行向量 $\boldsymbol{\mu}$ 的变差范数	(2.1.5)
$\\|f\\|_u$	函数 f 的一致范数	(2.1.10)
$\\|\boldsymbol{M}\\|_{u,v}$	矩阵 $\boldsymbol{M}$ 的一致变差范数	(3.2.1)
$\mathrm{Var}_{\mu}(f)$	f 关于概率向量 μ 的方差	

参考文献

1. Diaconis, P. & Stroock, D., *Geometric bounds for eigenvalues of Markov chains*, Ann. Appl. Probab. **1#1** (1991), 36-61.
2. Dunford, N. & Schwartz, J., *Linear Operators, part I*, Wiley Classics Lib., Wiley-Interscience, NY, 1988.
3. Holley, R. & Stroock, D., *Simulated annealing via Sobolev inequalities*, Comm. Math. Phys. **115#4** (1988), 553-569.
4. Karlin, S. & Taylor, H., *A First Course in Stochastic Processes, 2nd ed.*, Academic Press, NY, 1975.
5. Norris, J.R., *Markov Chains*, Cambridge Series in Statistical & Probabilistic Mathematics, Cambridge Univ. Press, Cambridge, U.K., 1997.
6. Revuz, D., *Markov Chains*, Mathematical Library, vol. 11, North Holland, Amsterdam & New York, 1984.
7. Riesz, F. & Sz.-Nagy, B., *Functional Analysis*, translated from the French edition by F. Boron, reprint of 1955 original, Dover Books, NY, 1990.
8. Stroock, D., *A Concise Introduction to the Theory of Integration, 3rd ed.*, Birkhäuser-Boston, Cambridge, MA, USA, 1998.
9. Stroock, D., *Probability Theory, An Analystic View, 2nd ed.*, Cambridge Univ. Press, NY, 2000.

索　引

Bernoulli 随机变量, 1
Borel σ-代数, 162
Cauchy 收敛, 28
Chapman-Kolmogorov 方程, 87
Dirichlet 型, 127
Doeblin 定理, 31
Doeblin 基本定理, 30
Doob h-变换, 46
Doob 停时定理, 52
Fatou 引理, 169
Gibbs 态, 138
Glauber 动力系统, 138
Gronwall 不等式, 137
Jensen 不等式, 152
Kolmogorov 向后方程, 93
Kolmogorov 向前方程, 94
Kronecker 符号 $\delta_{k,\ell}$, 3
Lebesgue 控制收敛定理, 169
Markov 不等式, 167
Markov 过程, 87
Markov 链, 25, 26
Markov 性, 25
Metropolies 算法, 143
Poincaré 常数, 128
Poincaré 不等式, 127
$\boldsymbol{P}$-平稳, 62
Q-矩阵, 94
$\boldsymbol{Q}$-互通, 104
$\boldsymbol{Q}$-不可约的, 104
$\boldsymbol{Q}$-瞬时, 105
$\boldsymbol{Q}$-正常返, 109
$\boldsymbol{Q}$-常返, 105
$\boldsymbol{Q}$-零常返, 109
Schwarz 不等式, 17
Stirling 公式, 20
σ-代数, 159
σ-代数的独立性, 178
σ-有限, 161

B

伴随概率矩阵, 118
半群, 87
爆炸时间, 98
变差范数, 27
遍历理论, 36
不可约的, 49

C

测度, 159
测度空间, 159
测度连续性, 161
常返的, 10
常返时刻, 9
常返性, 7
初始分布, 26
次可数可加性, 161

D

单调收敛定理, 169
独立随机变量, 178

E

二项系数, 2

F

反射路径, 4
反射原理, 4
范数, 28
非爆炸, 100
非负定, 127
非退化, 88
非周期的, 55
分布函数, 172, 180
分支过程, 44
符号函数, 5
复合 *Poisson* 过程, 85

G

概率, 172
概率测度, 172
概率空间, 172
概率向量, 26
更新方程, 61
广义二项系数, 6

H

互逆的, 43
互通类性质, 49
划分函数, 139

J

积分的单调性, 165
极点, 63
几乎处处, 167
简单 *Poisson* 过程, 82
简单函数, 165
经验测度, 80
卷积, 87

K

可测映射, 163
可测函数, 163
可测空间, 159
可测子集类, 159
可到达的, 39
可积函数, 166
可逆, 117
可逆 *Markov* 过程, 117
可容许的, 129
可数可加, 160

L

冷却方案, 144
零测度集, 167
零常返, 63

M

灭绝, 44

模拟退火, 143

N

内积, 119

O

耦合, 17

P

排队论, 23
排队模型, 23
平均遍历定理, 35
平稳分布, 30
瓶颈系数, 155
谱理论, 32

Q

期望, 172
齐次, 82
强大数律, 18

R

弱大数律, 17

S

三角不等式, 28
生成的 σ-代数, 162
时间 - 非齐次 *Kolmogorov* 向前方程, 144
时间 - 非齐次的转移概率, 144
事件, 172
首次返回的时刻, 7
首达时, 3
双随机, 43
瞬时的, 10
速率, 88
随机变量, 172
随机变量的分布, 172

T

条件概率, 2
条件期望, 180
凸函数, 151
凸集, 63

W

无穷小特征, 96

X

相互独立, 179
向后变量, 93
向前变量, 94

Y

样本点, 172
样本空间, 172
一致范数, 29
有界, 88

Z

正常返, 63
周期, 55
逐点遍历定理, 79
逐一平衡条件, 118, 132
转移概率矩阵, 26
状态空间, 25
自旋 - 反转 (*spin-flip*) 系统, 156
最大公约数, 55